OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2002 Compugen Ltd.

GenCore version 5.1.3

Run On: December 16, 2002, 10:54:58 ; Search time 68 Seconds

Title: US-09-435-471B-9
Perfect score: 33
Sequence: 1 gcattctagtcgacaagagtttgtcagtgggag 33

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19	57.6	2599	6	5266464	Patent No. 5266464
c 1	18.4	55.8	445	4	US-08-905-223-122	Sequence 122, App
c 2	18.4	55.8	445	4	US-08-905-223-120	Sequence 120, App
c 3	18.4	55.8	460	4	US-08-468-448-19	Sequence 19, Appl
c 4	18.4	55.8	2102	4	US-09-318-448-19	Sequence 19, Appl
c 5	18.2	55.2	2484	4	US-09-293-322C-4	Sequence 4, Appli
c 6	18	54.5	12847	1	US-08-550-715-1	Sequence 1, Appli
c 7	18	54.5	4169	1	US-09-166-350-32	Sequence 32, Appl
c 8	17.6	53.3	37	1	US-08-388-672A-5	Sequence 5, Appli
c 9	17.6	53.3	37	3	US-09-080-554-5	Sequence 5, Appli
c 10	17.6	53.3	150	2	US-08-943-731-85	Sequence 85, Appl
c 11	17.6	53.3	38682	4	US-08-943-731-2	Sequence 2, Appli
c 12	17.6	53.3	152331	3	US-09-128-155-16	Sequence 16, Appl
c 13	17.6	53.3	176373	3	US-09-128-155-17	Sequence 17, Appl
c 14	17.4	52.7	5261	1	US-09-045-806-3	Sequence 3, Appli
c 15	17.4	52.7	5261	1	US-08-366-051B-3	Sequence 3, Appli
c 16	17.4	52.7	72928	3	US-08-009-913-1	Sequence 1, Appli
c 17	17.2	52.1	2844	4	US-09-221-017B-415	Sequence 415, App
c 18	17	51.5	1614	4	US-09-134-078-9	Sequence 9, Appli
c 19	16.8	50.9	2429	4	US-09-360-545-68	Sequence 68, Appl
c 20	16.8	50.9	8855	2	US-08-542-003-1	Sequence 1, Appli
c 21	16.8	50.9	8855	2	US-08-322-760A-1	Sequence 1, Appli
c 22	16.8	50.9	8855	4	US-09-236-949-1	Sequence 1, Appli
c 23	16.6	50.3	230	3	US-08-974-180-10	Sequence 10, Appl
c 24	16.6	50.3	345	1	US-08-974-180-9	Sequence 9, Appli
c 25	16.6	50.3	640	4	US-09-221-017B-42	Sequence 42, Appl
c 26	16.6	50.3	2026	4	US-09-324-455-1	Sequence 1, Appli
c 27	16.6	50.3	2970	3	US-08-974-180-14	Sequence 14, Appl
28	16.6	50.3	3311	2	US-08-239-276-10	Sequence 10, Appl
29	16.6	50.3	3311	2	US-08-468-579B-10	Sequence 10, Appl
30	16.6	50.3	3311	3	US-08-468-579B-10	Sequence 10, Appl
31	16.6	50.3	3613	4	US-08-514-213A-1	Sequence 1, Appli
32	16.6	50.3	6476	4	US-09-127-670-5	Sequence 5, Appli
33	16.6	50.3	21234	4	US-09-810-671-3	Sequence 3, Appli
34	16.6	50.3	36741	4	US-09-301-665-3	Sequence 3, Appli
35	16.6	50.3	81001	4	US-09-750-580-1	Sequence 1, Appli
36	16.6	50.3	3600	4	PCT-US95-13749-1	Sequence 1, Appli
37	16.4	49.7	12047	2	US-09-022-461-1	Sequence 1, Appli
38	16.4	49.7	12047	2	US-09-033-556-3	Sequence 3, Appli
39	16.4	49.7	35100	2	US-08-770-379-19	Sequence 19, Appl
40	16.4	49.7	35100	4	US-08-757-669A-19	Sequence 19, Appl
41	16.4	49.7	35100	4	US-09-230-371A-19	Sequence 19, Appl
42	16.2	49.1	38	4	US-08-811-481-19	Sequence 19, Appl
43	16.2	49.1	682	1	US-08-463-115-21	Sequence 21, Appl
44	16.2	49.1	682	1	US-08-465-388-21	Sequence 21, Appl
45	16.2	49.1	1620	6	5449756-10	Patent No. 5449756

ALIGNMENTS

RESULT 1
;Patent No. 5266464
5266464-1/c
; APPLICANT: HOUSEY, GERARD
; TITLE OF INVENTION: METHOD OF SCREENING FOR PROTEIN INHIBITORS
; AND ACTIVATORS
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/392,073
; FILING DATE: 10-AUG-16989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 154,206
; FILING DATE: 10-FEB-1988
; SEQ ID NO:1:
; LENGTH: 2599

5266464-1

Query Match   57.6%;   Score 19;   DB 6;   Length 2599;
Best Local Similarity 100.0%;   Pred. No. 7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   15 AAGAGTTTGTCAGTGGGAG 33
           |||||||||||||||||||
Db   2034 AAGAGTTTGTCAGTGGGAG 2016

RESULT 2
US-08-905-223-122
; Sequence 122, Application US/08905223
; Patent NO. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223

FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 122:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 236..418
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 4.6
OTHER INFORMATION: seq LILLHGGGHSALS/WA
US-08-905-223-122

```
Qy    4 TTCTAGTCGACAAGAGTTTGTCAGTGGG  31
          ||| |||| |||||||| |||||| |||
Db  345 TTCGAGTCTACAAGAGTGGTTCAGAGGG 372
```

Query Match   55.8%;   Score 18.4;   DB 4;   Length 445;
Best Local Similarity   78.6%;   Pred. No. 9.1;
Matches 22;   Conservative 0;   Mismatches 6;   Indels 0;   Gaps 0;

US-08-905-223-120
RESULT 3
Sequence 120, Application US/08905223
Patent No. 6222029
GENERAL INFORMATION:
APPLICANT: Edwards, Jean-Baptiste D.
APPLICANT: Duclert, Aymeric
APPLICANT: Lacroix, Bruno
TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
NUMBER OF SEQUENCES: 503
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
ATTORNEY/AGENT INFORMATION:
CLASSIFICATION: 536
NAME: Israelsen, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 460 base pairs
TYPE: NUCLEIC ACID

STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 254..436
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 4.6
OTHER INFORMATION: seq LILLHGGGHSALS/WA
US-08-905-223-120

```
Qy    4 TTCTAGTCGACAAGAGTTTGTCAGTGGG  31
          ||| |||| |||||||| |||||| |||
Db  363 TTCGAGTCTACAAGAGTGGTTCAGAGGG 390
```

Query Match   55.8%;   Score 18.4;   DB 4;   Length 460;
Best Local Similarity   78.6%;   Pred. No. 9.2;   Mismatches 6;   Indels 0;   Gaps 0;
Matches 22;   Conservative 0;

RESULT 4
US-09-318-448-19/c
Sequence 19, Application US/09318448
Patent No. 6210950
GENERAL INFORMATION:
APPLICANT: Johnson, William G.
APPLICANT: Stenroos, Edward S.
TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
FILE REFERENCE: 601-1-057
CURRENT APPLICATION NUMBER: US/09/318,448
CURRENT FILING DATE: 1999-05-25
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 19
LENGTH: 2102
TYPE: DNA
ORGANISM: Homo sapiens
US-09-318-448-19

```
Qy     5 TCTAGTCGACAAGAGTTTGTCAGTGGGA   32
           ||| |||| ||||||| ||||||||||
Db  1606 TCCAGCAGGCAAGAGTAGGTCAGTGGGA 1579
```

Query Match   55.8%;   Score 18.4;   DB 4;   Length 2102;
Best Local Similarity   78.6%;   Pred. No. 13;   Mismatches 6;   Indels 0;   Gaps 0;
Matches 22;   Conservative 0;

RESULT 5
US-09-293-322C-4
Sequence 4, Application US/09293322C
Patent No. 6233110
GENERAL INFORMATION:
APPLICANT: Pallas, David C
APPLICANT: Du, Xianing
TITLE OF INVENTION: Coding Sequence for Protein Phosphatase Methylesterase,
TITLE OF INVENTION: Recombinant DNA Molecules and Methods
Patent No. 6233110
FILE REFERENCE: I05-97
CURRENT APPLICATION NUMBER: US/09/293,322C
CURRENT FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: US 60/082,202
PRIOR FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 2484
TYPE: DNA
ORGANISM: Homo sapiens

```
;  FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (100)..(1257)
   US-09-293-322C-4

Qy    4 TTCTAGTCGACAAGAGTTTGTCAGTGGG 31
        ||| |||| |||||||| |||| |||
Db  299 TTCGAGTCTACAAGAGTGGGTTCAGAGGG 326

Query Match          55.8%;  Score 18.4;  DB 4;  Length 2484;
Best Local Similarity 78.6%;  Pred. No. 13;
Matches  22;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

RESULT 6
US-08-550-715-1/c
Sequence 1, Application US/08550715
Patent No. 5750345
GENERAL INFORMATION:
   APPLICANT: Bowie, Lemuel J.
   TITLE OF INVENTION: Human \-Thalassemia Mutations as a Predictor of
   TITLE OF INVENTION: Blood-Related Disorders
   NUMBER OF SEQUENCES: 13
   CORRESPONDENCE ADDRESS:
     ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
     STREET: 6300 Sears Tower, 233 South Wacker Drive
     CITY: Chicago
     STATE: Illinois
     COUNTRY: United States of America
     ZIP: 60606-6402
   COMPUTER READABLE FORM:
     MEDIUM TYPE: Floppy disk
     COMPUTER: IBM PC compatible
     OPERATING SYSTEM: PC-DOS/MS-DOS
     SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER: US/08/550,715
     FILING DATE:
   CLASSIFICATION: 435
   ATTORNEY/AGENT INFORMATION:
     NAME: Gass, David A.
     REGISTRATION NUMBER: 38,153
     REFERENCE/DOCKET NUMBER: 28493/32834
   TELECOMMUNICATION INFORMATION:
     TELEPHONE: 312/474-6300
     TELEFAX: 312/474-0448
     TELEX: 25-3856
   INFORMATION FOR SEQ ID NO: 1:
     SEQUENCE CHARACTERISTICS:
       LENGTH: 12847 base pairs
       TYPE: nucleic acid
       STRANDEDNESS: single
       TOPOLOGY: linear
     MOLECULE TYPE: DNA
     FEATURE:
       NAME/KEY: CDS
       LOCATION: join(6703..6797, 6915..7119, 7262..7387)
   US-08-550-715-1
     FEATURE:
       NAME/KEY: CDS
       LOCATION: join(10514..10608, 10726..10930, 11080..11205)

Qy    1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGG 31
        |||||| || || ||| ||| |||| |||
Db 9463 GCATTCAAGGGACCAGGGTTAGTCTCGAGGG 9433

RESULT 7
US-09-166-350-32/c
```

```
;  Sequence 32, Application US/09166350A
;  Patent NO. 6440663
;  GENERAL INFORMATION:
;    APPLICANT: Chen, Yao
;    APPLICANT: Scanlan, Matthew
;    APPLICANT: Stockert, Elisabeth
;    APPLICANT: Old, Lloyd
;    APPLICANT: Jager, Elke
;    APPLICANT: Knuth, Alex
;    TITLE OF INVENTION: Renal Cancer Associated Antigens and
;    TITLE OF INVENTION: Uses Therefor
;    FILE REFERENCE: L0461/7051
;    CURRENT APPLICATION NUMBER: US/09/166,350
;    CURRENT FILING DATE: 1998-10-05
;    EARLIER APPLICATION NUMBER: US 09/166,350
;    EARLIER FILING DATE: 1998-10-05
;    NUMBER OF SEQ ID NOS: 35
;    SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 32
;    LENGTH: 4169
;    TYPE: DNA
;    ORGANISM: Homo sapiens
   US-09-166-350-32

Qy    4 TTCTAGTCGACAAGAGTTTGTCAGTG 29
        ||||| ||| ||||||||||||||||
Db  940 TTATAGTAAACATGAATTTGTCAGTG 915

Query Match          54.5%;  Score 18;  DB 4;  Length 4169;
Best Local Similarity 80.8%;  Pred. No. 23;
Matches  21;  Conservative  0;  Mismatches  5;  Indels  0;  Gaps  0;

RESULT 8
US-08-388-672A-5
Sequence 5, Application US/08388672A
Patent NO. 5795961
GENERAL INFORMATION:
   APPLICANT: Wallace, T. Paul
   APPLICANT: Harris, William J.
   APPLICANT: Carr, Frank J.
   APPLICANT: Old, Lloyd J.
   APPLICANT: Welt, Sydney
   APPLICANT: Kitamura, Kunio
   TITLE OF INVENTION: Recombinant Human Anti-Lewis B
   TITLE OF INVENTION: Antibodies
   NUMBER OF SEQUENCES: 25
   CORRESPONDENCE ADDRESS:
     ADDRESSEE: Felfe and Lynch
     STREET: 805 Third Avenue
     CITY: New York
     STATE: New York
     COUNTRY: U.S.A.
     ZIP: 10022
   COMPUTER READABLE FORM:
     MEDIUM TYPE: Floppy disk
     COMPUTER: IBM PC compatible
     OPERATING SYSTEM: PC-DOS/MS-DOS
     SOFTWARE: PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER: US/08/388,672A
     FILING DATE: 14-FEB-1995
   CLASSIFICATION:
   ATTORNEY/AGENT INFORMATION:
     NAME: Hanson, No. 5795961man D.
     REGISTRATION NUMBER: 30,946
     REFERENCE/DOCKET NUMBER: LUD 5409
   TELECOMMUNICATION INFORMATION:
     TELEPHONE: 212-688-9200
     TELEFAX: 212-838-3884
   INFORMATION FOR SEQ ID NO: 5:
     SEQUENCE CHARACTERISTICS:
       LENGTH: 37 base pairs
```

```
;       TYPE: nucleic acid
;       STRANDEDNESS: unknown
;       TOPOLOGY: unknown
;       MOLECULE TYPE: DNA (genomic)
US-08-388-672A-5

Query Match      53.3%;  Score 17.6;  DB 1;  Length 37;
Best Local Similarity   65.4%;  Pred. No. 13;
Matches  17;  Conservative  5;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    6 CTAGTCGCACAAGAGAGTTTGTCAGTGGG 31
        |||||||||| ||| |:|:: |:| :|
Db    2 CTAGTCGACAATGAGAGGTKGYYTGYTSG 27
```

RESULT 9
US-09-080-554-5
; Sequence 5, Application US/09080554
; Patent No. 6025481
; GENERAL INFORMATION:
;   APPLICANT: Wallace, T Paul
;   APPLICANT: Harris, William J.
;   APPLICANT: Carr, Frank J.
;   APPLICANT: Old, Lloyd J.
;   APPLICANT: Welt, Sydney
;   APPLICANT: Kitamura, Kunio
;   TITLE OF INVENTION: Recombinant Human Anti-Lewis B
;   TITLE OF INVENTION: Antibodies
;   NUMBER OF SEQUENCES: 25
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Pennie & Edmonds
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/080,554
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/388,672
;   FILING DATE: 14-FEB-1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Halluin, Albert P.
;   REGISTRATION NUMBER: 25,227
;   REFERENCE/DOCKET NUMBER: 7606-032
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-854-3660
;   TELEFAX: 415-854-3694
;   TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 37 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA (genomic)
US-09-080-554-5

```
Query Match      53.3%;  Score 17.6;  DB 3;  Length 37;
Best Local Similarity   65.4%;  Pred. No. 13;
Matches  17;  Conservative  5;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    6 CTAGTCGCACAAGAGTTTGTCAGTGGG 31
        |||||||||| ||| |:|:: |:| :|
Db    2 CTAGTCGACAATGAGGTKGYYTGYTSG 27
```

RESULT 10
US-08-943-731-85/c
; Sequence 85, Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
;   APPLICANT: PROKOP, DARWIN J.
;   APPLICANT: SPOTILA, LORETTA D.
;   APPLICANT: DELTAS, CONSTANTINOS D.
;   APPLICANT: SEREDA, LARISA
;   APPLICANT: LARSON, ANDREA W.
;   APPLICANT: PACK, MICHAEL
;   APPLICANT: COLIGE, ALAIN
;   APPLICANT: EARLY, JAMES
;   APPLICANT: KORKKO, JARMO
;   APPLICANT: ALA-KOKKO, LEENA, et al.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
;   TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
;   NUMBER OF SEQUENCES: 666
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
;   STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;   CITY: PHILADELPHIA
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19103-7086
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/943,731
;   FILING DATE: 03-OCT-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/212,322
;   FILING DATE: 14-MAR-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/803,628
;   FILING DATE: 03-DEC-1991
;   ATTORNEY/AGENT INFORMATION:
;   NAME: DOYLE LEARY Ph.D., KATHRYN
;   REGISTRATION NUMBER: 36,317
;   REFERENCE/DOCKET NUMBER: 9598-27
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-965-1284
;   TELEFAX: 215-567-2991
;   TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 85:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 150 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-943-731-85

```
Query Match      53.3%;  Score 17.6;  DB 4;  Length 150;
Best Local Similarity   71.9%;  Pred. No. 17;
Matches  23;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy    2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGAG 33
        ||||||||| | ||||| ||||||||
Db   42 CATTCTAGTCGAGAAGAGCACACAAGAGCTGGGAG 11
```

RESULT 11
US-08-943-731-2/c
; Sequence 2, Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:

; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DELTAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISA
; APPLICANT: LARSON, ANDREA W.
; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORKKO, JARMO
; APPLICANT: ALA-KOKKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
;   STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
;   CITY: PHILADELPHIA
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19103-7086
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/943,731
;   FILING DATE: 03-OCT-1997
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/212,322
;   FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/803,628
;   FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: DOYLE LEARY Ph.D., KATHRYN
;   REGISTRATION NUMBER: 36,317
;   REFERENCE/DOCKET NUMBER: 9598-27
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 215-965-1284
;   TELEFAX: 215-567-2991
;   TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 38682 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)

Query Match   53.8%; Score 17.6; DB 4; Length 38682;
Best Local Similarity 71.9%; Pred. No. 58;
Matches   23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

US-08-943-731-2

Qy   2 CATTCTAGTCGACAAGAGAGTTTGTCAGTGGGAG 33
          ||||||||| |||||| |||||||
Db 21154 CATTCTAGTCGAGAGAGCACAAGAGAGTGGGAG 21123

RESULT 12
US-09-128-155-16/C
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;   APPLICANT: Pan, Yang
;   TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;   TITLE OF INVENTION: AND USES THEREOF
;   FILE REFERENCE: 0404/052001
;   CURRENT APPLICATION NUMBER: 09/128,155
;   CURRENT FILING DATE: 1998-08-03

; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match   53.8%; Score 17.6; DB 3; Length 152331;
Best Local Similarity 71.9%; Pred. No. 77;
Matches   23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy   1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
          ||| |||| | |||||||| |||||| ||||
Db 115561 GCAAGCCAAGTGACAAGAGTTTTGTCCCTGGGA 115530

RESULT 13
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;   APPLICANT: Pan, Yang
;   TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;   TITLE OF INVENTION: AND USES THEREOF
;   FILE REFERENCE: 0404/052001
;   CURRENT APPLICATION NUMBER: US/09/128,155
;   CURRENT FILING DATE: 1998-08-03
;   EARLIER APPLICATION NUMBER: US 60/091,650
;   EARLIER FILING DATE: 1998-07-02
;   EARLIER APPLICATION NUMBER: US 60/054,646
;   EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match   53.8%; Score 17.6; DB 3; Length 176373;
Best Local Similarity 71.9%; Pred. No. 79;
Matches   23; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy   1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
          ||| |||| | |||||||| |||||| ||||
Db 63844 GCAAGCCAAGTGACAAGAGTTTTGTCCCTGGGA 63875

RESULT 14
US-08-045-806-3
; Patent No. 5378822
; GENERAL INFORMATION:
;   APPLICANT: Bradfield, Christopher Alan
;   APPLICANT: Dolwick, Kristin Marie
;   APPLICANT: Poland, Alan
;   TITLE OF INVENTION: Ah Receptor cDNA and Method of
;   TITLE OF INVENTION: Determining Human Risks To Environmental Pollutants
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:

ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
STREET: 100 South Wacker Drive, Suite 960
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-4002
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/045,806
FILING DATE: 19930408
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fentress, Susan B.
REGISTRATION NUMBER: 31,327
REFERENCE/DOCKET NUMBER: NU-9207
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)-456-8000
TELEFAX: (312)-456-7776
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5261 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 383..2927
US-08-045-806-3

Query Match
Best Local Similarity   52.7%;   Score 17.4;   DB 1;   Length 5261;
Matches 21;   Conservative   77.8%;   Pred. No. 47;
0;   Mismatches   6;   Indels 0;   Gaps 0;

Qy   4 TTCTAGTCGACAAGAGTTTGTCAGTGG 30
          |||||||| | ||| |||||||| |
Db 2578 TTCTAGTTTAGAAGATTTGTCACTTG 2604

RESULT 15
US-08-366-051B-3
Sequence 3, Application US/08356051B
Patent No. 5650283
GENERAL INFORMATION:
APPLICANT: Bradfield, Christopher A.
APPLICANT: Dolwick, Kristin M.
APPLICANT: Carver, Lucy A.
TITLE OF INVENTION: Ah Receptor cDNAs and Genetically
TITLE OF INVENTION: Engineered Cells for Detecting Agonists to the Ah
TITLE OF INVENTION: Receptor
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
STREET: 100 South Wacker Drive, Suite 960
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-4002
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,051B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tilton, Timothy L.

REGISTRATION NUMBER: 16,926
REFERENCE/DOCKET NUMBER: NU-9207-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)-456-8000
TELEFAX: (312)-456-7776
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5261 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 383..2927
US-08-366-051B-3

Query Match
Best Local Similarity   52.7%;   Score 17.4;   DB 1;   Length 5261;
Matches 21;   Conservative   77.8%;   Pred. No. 47;
0;   Mismatches   6;   Indels 0;   Gaps 0;

Qy   4 TTCTAGTCGACAAGAGTTTGTCAGTGG 30
          |||||||| | ||| |||||||| |
Db 2578 TTCTAGTTTAGAAGATTTGTCACTTG 2604

Search completed: December 16, 2002, 10:57:22
Job time : 128 secs

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2002 Compugen Ltd.

GenCore version 5.1.3

Run on: December 16, 2002, 10:54:59 ; Search time 2429 seconds
              (without alignments)
              220.029 Million cell updates/sec

Title: US-09-435-471B-9
Perfect score: 33
Sequence: 1 gcattctagtcgacaagagtttgtcagtgggag 33
Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues
Total number of hits satisfying chosen parameters: 32308132
Maximum DB seq length: 2000000000
Minimum DB seq length: 0
Post-processing: Minimum Match 0%
              Maximum Match 100%
              Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estom:*
16: em_estfun:*
17: gb_gss:*
18: em_gss:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21.4	64.8	1061	17	BH677152	BH677152 BOMKA35TR
2	21.2	64.2	788	12	BG780076	BG780076 SEAUMC000
3	20.8	63.0	427	17	BH252660	BH252660 SALK_0137
4	20.8	63.0	1029	10	BE563270	BE563270 601335580
5	20.6	62.4	698	17	AZ613804	AZ613804 1M0442017
6	20.4	61.8	660	17	AG116222	AG116222 Pan trogl

	No.	Score	% Query Match	Length	DB	ID	Description
c	7	20.4	61.8	766	12	BG614194	BG614194 602641866
c	8	20.2	61.2	252	9	AA359569	AA359569 EST68605
c	9	20.2	61.2	403	14	D26831	D26831 CELK09E2F
c	10	20	60.6	466	12	BF851402	BF851402 IL5-EN008
c	11	20	60.6	697	13	BI224033	BI224033 602429940
c	12	20	60.6	728	17	BH598122	BH598122 BOGL253TF
c	13	20	60.6	833	17	BH449843	BH449843 BOGRE19TR
c	14	20	60.6	978	13	BM468802	BM468802 AGENCOURT
c	15	20	60.6	240	10	BB150049	BB150049 BB150049
c	16	19.8	60.0	281	10	BB159127	BB159127 BB159127
c	17	19.8	60.0	444	12	BF490151	BF490151 AT26449.5
c	18	19.8	60.0	611	17	AZ832714	AZ832714 2M0113K10
c	19	19.8	60.0	886	12	BF030538	BF030538 601559867
c	20	19.8	60.0	940	17	AG077093	AG077093 Pan trogl
c	21	19.6	59.4	219	12	BEB27967	BEB27967 RC4-ET002
c	22	19.6	59.4	401	13	BM448487	BM448487 DSA01LF07
c	23	19.6	59.4	407	12	BF996133	BF996133 RC3-GNO04
c	24	19.6	59.4	586	9	AU199787	AU199787 AU199787
c	25	19.6	59.4	615	13	BJ120449	BJ120449 BJ120449
c	26	19.6	59.4	770	13	BI824720	BI824720 603033888
c	27	19.4	58.8	183	10	AV785600	AV785600 AV785600
c	28	19.4	58.8	267	17	BB576883	BB576883 BB576883
c	29	19.4	58.8	413	10	BE068204	BE068204 CM1-BT036
c	30	19.4	58.8	556	17	AQ113920	AQ113920 CIT-HSP-2
c	31	19.4	58.8	620	10	A2804633	A2804633 2M0065M10
c	32	19.4	58.8	329	17	AW761254	AW761254 s165c01.y
c	33	19.2	58.2	329	10	AA874978	AA874978 UI-R-E0-c
c	34	19.2	58.2	235	10	AI452286	AI452286 mp90c02.x
c	35	19.2	58.2	345	9	AU110888	AU110888 AU110888
c	36	19.2	58.2	386	9	AA719026	AA719026 ah46c01.s
c	37	19.2	58.2	390	12	BF440482	BF440482 BS290000I4
c	38	19.2	58.2	424	17	AQ845191	AQ845191 am44e08_J
c	39	19.2	58.2	425	9	AA451381	AA451381 vf81a03.r
c	40	19.2	58.2	461	9	AI045963	AI045963 UI-R-C1-j
c	41	19.2	58.2	489	13	BJ481882	BJ481882 BJ481882
c	42	19.2	58.2	507	13	BJ292333	BJ292333 BJ292333
c	43	19.2	58.2	514	12	BF630170	BF630170 HVSMEB000
c	44	19.2	58.2	529	9	AA690593	AA690593 vu53d05.r
c	45	19.2	58.2	540	9	AJ470270	AJ470270 AJ470270

RESULT 1

BH677152/c

```
LOCUS       BH677152              1061 bp    DNA    linear   GSS 19-FEB-2002
DEFINITION  BOMKA35TR BO_2_3_KB Brassica oleracea genomic clone BOMKA35, DNA
            sequence.
ACCESSION   BH677152
VERSION     BH677152.1  GI:18747595
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 1061)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..1061
                     /organism="Brassica oleracea"
```

ALIGNMENTS

	ID	Description
c	BG614194	BG614194 602641866
c	AA359569	AA359569 EST68605
c	D26831	D26831 CELK09E2F
c	BF851402	BF851402 IL5-EN008
c	BI224033	BI224033 602429940
c	BH598122	BH598122 BOGL253TF
c	BH449843	BH449843 BOGRE19TR
c	BM468802	BM468802 AGENCOURT
c	BB150049	BB150049 BB150049
c	BB159127	BB159127 BB159127
c	BF490151	BF490151 AT26449.5
c	AZ832714	AZ832714 2M0113K10
c	BF030538	BF030538 601559867
c	AG077093	AG077093 Pan trogl
c	BEB27967	BEB27967 RC4-ET002
c	BM448487	BM448487 DSA01LF07
c	BF996133	BF996133 RC3-GNO04
c	AU199787	AU199787 AU199787
c	BJ120449	BJ120449 BJ120449
c	BI824720	BI824720 603033888
c	AV785600	AV785600 AV785600
c	BB576883	BB576883 BB576883
c	BE068204	BE068204 CM1-BT036
c	AQ113920	AQ113920 CIT-HSP-2
c	A2804633	A2804633 2M0065M10
c	AW761254	AW761254 s165c01.y
c	AA874978	AA874978 UI-R-E0-c
c	AI452286	AI452286 mp90c02.x
c	AU110888	AU110888 AU110888
c	AA719026	AA719026 ah46c01.s
c	BF440482	BF440482 BS290000I4
c	AQ845191	AQ845191 am44e08_J
c	AA451381	AA451381 vf81a03.r
c	AI045963	AI045963 UI-R-C1-j
c	BJ481882	BJ481882 BJ481882
c	BJ292333	BJ292333 BJ292333
c	BF630170	BF630170 HVSMEB000
c	AA690593	AA690593 vu53d05.r
c	AJ470270	AJ470270 AJ470270

BASE COUNT   267 a   267 c   242 g   285 t
ORIGIN
       /strain="TO100DH3"
       /db_xref="taxon:3712"
       /clone="BONKA35"
       /clone_lib="BO_2_3_KB"
       /note="vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
       genomic DNA inserted into pHOS1 using BstXI linkers"

Query Match   64.8%;   Score   21.4;   DB 17;   Length 1061;
Best Local Similarity   80.6%;   Pred. No. 97;
Matches   25;   Conservative   0;   Mismatches   6;   Indels   0;   Gaps   0;

Qy   3 ATTCTAGTCGACAAGAGAGTTTGTCAGTGGGAG 33
        ||||| | |||||||||||||||||||||| |||
Db 342 ATTTTATTCATCAAGAGAGTTTGTCACTGAGAG 312

RESULT 2
LOCUS   BG780076   788 bp   mRNA   linear   EST 20-MAY-2001
DEFINITION BG780076 SEAUMC000033 Sea urchin primary mesenchyme cell cDNA
            Strongylocentrotus purpuratus cDNA clone 91214953_PC_0022_B2_C09_MR
            5', mRNA sequence.
ACCESSION   BG780076
VERSION   BG780076.1 GI:14151089
KEYWORDS   EST.
SOURCE   Strongylocentrotus purpuratus.
  ORGANISM Strongylocentrotus purpuratus.
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1 (bases 1 to 788)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE   A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
  MEDLINE   21384984
COMMENT   Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.
FEATURES   Location/Qualifiers
     source   1..788
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="91214953_PC_0022_B2_C09_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT   210 a   143 c   159 g   276 t
ORIGIN

Query Match   64.2%;   Score   21.2;   DB 12;   Length 788;
Best Local Similarity   88.5%;   Pred. No. 1e+02;
Matches   23;   Conservative   0;   Mismatches   3;   Indels   0;   Gaps   0;

Qy   4 TTCTAGTCGACAAGAGTTTGTCAGTG 29
        |||| |||||||| ||||||||||| |
Db 727 TTCTTGTCGACATGAGTTTGTCAGCG 752

RESULT 3
                    BH252660
LOCUS   BH252660/c
DEFINITION SALK_013711 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_013711, DNA sequence.
ACCESSION   BH252660
VERSION   BH252660.1 GI:17139638
KEYWORDS   GSS.
SOURCE   thale cress.
            Arabidopsis thaliana
  ORGANISM Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 427)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE   Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Joseph R. Ecker
            A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES   Location/Qualifiers
     source   1..427
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_013711"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains one or more TDNA insertion
                     elements. The resultant fragment for each line was
                     directly sequenced to determine the genomic sequence at
                     the site of insertion. Details of the protocols used can
                     be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT   110 a   111 c   87 g   111 t
ORIGIN

Query Match   63.0%;   Score   20.8;   DB 17;   Length 427;
Best Local Similarity   78.1%;   Pred. No. 1.1e+02;
Matches   25;   Conservative   0;   Mismatches   7;   Indels   0;   Gaps   0;

Qy   1 GCATTCTAGTCGACAAGAGAGTTTGTCAGTGGGA 32
        | ||||||||| |||||| || | | ||||| ||
Db   63 GAATTCTAGTCAACAAGAGGGCGTAAGTGGGTA 32

RESULT 4
                    BE563270
LOCUS   BE563270   1029 bp   mRNA   linear   EST 15-AUG-2000
DEFINITION 601355580F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689707 5',
            mRNA sequence.
ACCESSION   BE563270
VERSION   BE563270.1 GI:9807082
KEYWORDS   EST.
SOURCE   human.
            Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1029)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM384 row: o column: 20
High quality sequence stop: 669.

FEATURES
  source   Location/Qualifiers
                  1..1029
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3689707"
                  /clone_lib="NIH_MGC_39"
                  /tissue_type="adenocarcinoma"
                  /lab_host="DH10B (phage-resistant)"
                  /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                  Site_2: EcoRI; cDNA made by oligo-dT priming.
                  Directionally cloned into EcoRI/XhoI sites using the
                  following 5' adaptor: GGCACGAG(G). Library constructed
                  by Ling Hong in the laboratory of Gerald M. Rubin
                  (University of California, Berkeley) using ZAP-cDNA
                  synthesis kit (Stratagene) and Superscript II RT (Life
                  Technologies)."

BASE COUNT   314 a   232 c   302 g   181 t
ORIGIN

Qy   8 AGTCGCACAAGAGTTTGTCAGTGGG 31
          ||||||||||||||||||||||||||
Db   763 AATTGACAAGAGTTTGTCAGTGGG 786

Query Match   63.0%; Score 20.8; DB 10; Length 1029;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

RESULT 5
LOCUS   AZ613804   698 bp   DNA   linear   GSS 13-DEC-2000
DEFINITION IM0442017F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0442017 F, DNA sequence.
ACCESSION   AZ613804
VERSION   AZ613804.1 GI:11735994
KEYWORDS   GSS.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 698)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
  TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0442 row: O column: 17
            Seq primer: CGTGTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 698.

FEATURES
  source   Location/Qualifiers
                  1..698
                  /organism="Mus musculus"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="UUGC1M0442017"
                  /clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi|473214|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT   198 a   150 c   161 g   189 t
ORIGIN

Qy   3 ATTCTAGTCGCACAAGAGTTTGTCAGTG 29
          |||||||| | ||||||| |||||||| |
Db   414 ATTCTAGCCAACAAGAATTTGTCAGGG 388

Query Match   62.4%; Score 20.6; DB 17; Length 698;
Best Local Similarity 85.2%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

RESULT 6
LOCUS   AG116222   660 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-123L03.R, genomic survey sequence.
ACCESSION   AG116222
VERSION   AG116222.1 GI:16736741
KEYWORDS   GSS.
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB chimpanzee Male
            BAC Library clone:PTB-123L03.R.
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 660)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
                   Sequencing: M13Rev
            LIBRARY
                   Vector   : pKS145
                   R.Site 1   : SacI
                   R.Site 2   : SacI

FEATURES
  source   Location/Qualifiers
                  1..660
                  /organism="Pan troglodytes"
                  /db_xref="taxon:9598"
                  /clone="PTB-123L03.R"

Left column

```
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT        187 a      135 c      126 g      212 t
ORIGIN

Query Match
Best Local Similarity  61.8%;  Score 20.4;  DB 17;  Length 660;
Matches  24;  Conservative   0;  Pred. No. 2.1e+02;  Mismatches  6;  Indels  0;  Gaps  0;

Qy    3 ATTCTAGTCGACAAAGAGTTTGTCAGTGGGA 32
        ||||| ||| || | |||||||||||||||| |
Db  564 ATTCTTGTCAAGGAAAGTTTGTCAGTGGGAA 593
```

```
RESULT 7
LOCUS       BG614194                766 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602641866F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4772705 5',
            mRNA sequence.
ACCESSION   BG614194
VERSION     BG614194.1  GI:13665565
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 766)
  AUTHORS   NIH-MGC  http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1643  row: d  column: 18
            High quality sequence stop: 648.
FEATURES             Location/Qualifiers
     source          1..766
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4772705"
                     /clone_lib="NIH_MGC_61"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CAGGCGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGCGGCCGCGGGCCGCGGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     full-length clones and was enriched for
                     Library."
BASE COUNT        214 a      169 c      143 g      240 t
ORIGIN

Query Match
Best Local Similarity  61.8%;  Score 20.4;  DB 12;  Length 766;
Matches  24;  Conservative   0;  Pred. No. 2.3e+02;  Mismatches  6;  Indels  0;  Gaps  0;

Qy    3 ATTCTAGTCGACAAAGAGTTTGTCAGTGGGA 32
        ||||| ||| || | |||||||||||||||| |
Db  279 ATTCTTGTCAAGGAAAGTTTGTCAGTGGGAA 308
```

Right column

```
RESULT 8
LOCUS       AA359569                252 bp    mRNA    linear   EST 21-APR-1997
DEFINITION  EST68605 Fetal lung II Homo sapiens cDNA clone 5' end, mRNA sequence.
ACCESSION   AA359569
VERSION     AA359569.1  GI:2011885
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 252)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.,
            Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC77447
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..252
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):163683"
                     /db_xref="taxon:9606"
                     /clone_lib="Fetal lung II"
                     /sex="male"
                     /dev_stage="fetus, 19 wks"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI"
BASE COUNT         45 a       86 c       67 g       51 t      3 others
ORIGIN

Query Match
Best Local Similarity  61.2%;  Score 20.2;  DB 9;  Length 252;
Matches  25;  Conservative   0;  Pred. No. 1.5e+02;  Mismatches  8;  Indels  0;  Gaps  0;

Qy    1 GCATTCTAGTCGACAAAGAGTTTGTCAGTGGGAG 33
        || |||||| | ||||||||| ||| | |||||
Db  209 GCCTTCTATTCTGCAAGTTGTCAAGTTGACTTGGGAG 241
```

```
RESULT 9
LOCUS       D26831
DEFINITION  CELK009EZF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk9e12 5', mRNA sequence.
```

```
ACCESSION   D26831
VERSION     D26831.1  GI:521769
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina; Rhabditomorpha; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 403)
AUTHORS     Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
TITLE       Toward an expression map of the C.elegans genome
JOURNAL     Unpublished (1994)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1..403
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk9e12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT    133 a     78 c     88 g    102 t      2 others
ORIGIN

Qy        3 ATTCTAGTCGACAAGAGTTTGTCAGTGG 30
            |||||| || |||||| |||||| |||||
Db      168 ATTCTGGTTGACAAGCTTTTGTGAGTGG 141

Query Match   60.6%;  Score 20;  DB 14;  Length 403;
Best Local Similarity 82.1%;  Pred. No. 2.4e+02;
Matches 23;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
RESULT 10
LOCUS       BF851402                 466 bp    mRNA    linear   EST 16-JAN-2001
DEFINITION  IL5-EN0086-281100-282-f08 EN0086 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF851402
VERSION     BF851402.1  GI:12238564
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 466)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE       Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE     20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2704701
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL5&t2=IL5-EN0086-
            281100-282-f08&t3=2000-11-28&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 8
            High quality sequence stop: 396.
FEATURES             Location/Qualifiers
     source          1..466
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0086"
                     /dev_stage="Adult"
                     /note="Organ: lung normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the puc 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT     98 a    132 c    123 g    112 t      1 others
ORIGIN

Qy        5 TCTAGTCGACAAGAGTTTGTCAGTGGGA 32
            ||||| || |||||| |||||||||||||
Db      358 TCTAGCAGGCAAGAGTAGGTCAGTGGGA 331

Query Match   60.6%;  Score 20;  DB 12;  Length 466;
Best Local Similarity 82.1%;  Pred. No. 2.6e+02;
Matches 23;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
RESULT 11
LOCUS       BI224033                 697 bp    mRNA    linear   EST 11-JUL-2001
DEFINITION  602943940F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5106272 5',
            mRNA sequence.
ACCESSION   BI224033
VERSION     BI224033.1  GI:14677477
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 697)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1256 row: o column: 09
            High quality sequence stop: 614.
FEATURES             Location/Qualifiers
     source          1..697
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5106272"
                     /clone_lib="NIH_MGC_12"
                     /tissue_type="cervical carcinoma cell line"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.4 kb. Library prepared by Life
                     Technologies."
BASE COUNT    138 a    197 c    180 g    182 t
ORIGIN

Query Match   60.6%;  Score 20;  DB 13;  Length 697;
```

RESULT 12

Qy   6 CTAGTCGACAAGAGTTTGTCAGTGGGAG 33
        || || |||||||||| || ||| |||||
Db 501 CTCGTGGACAAGAGTATGCCAGAGGAG 528

Best Local Similarity   82.1%; Pred. No. 3.2e+02;
Matches 23; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

LOCUS   BH598122   728 bp   DNA   linear   GSS 15-DEC-2001
DEFINITION BOGL253TF BOGL Brassica oleracea genomic clone BOGL253, DNA
            sequence.
ACCESSION   BH598122
VERSION   BH598122.1   GI:17850574
KEYWORDS   GSS.
SOURCE   Brassica oleracea
  ORGANISM Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 728)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE   Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT   Other_GSSs: BOGL253TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            Seq primer: TF
            Class: sheared ends.
            DNA is from a doubled haploid provided by Tom Osborn.
FEATURES   Location/Qualifiers
     source   1..728
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGL253"
                     /clone_lib="BOGL"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT   211 a   171 c   158 g   188 t
ORIGIN

RESULT 13

Qy   3 ATTCTAGTCGACAAGAGTTTGTCAGTGG 30
        | ||||| |||||||| ||| ||||||||
Db 615 AATCTAGCCGTCAAGTGTCGGTCAGTGG 642

Query Match   60.6%; Score 20; DB 17; Length 833;
Best Local Similarity   82.1%; Pred. No. 3.3e+02;
Matches 23; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

LOCUS   BH449843   833 bp   DNA   linear   GSS 12-DEC-2001
DEFINITION BOGRE19TR BOGR Brassica oleracea genomic clone BOGRE19, DNA
            sequence.
ACCESSION   BH449843
VERSION   BH449843.1   GI:17635554
KEYWORDS   GSS.
SOURCE   Brassica oleracea
  ORGANISM Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 833)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE   Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)

COMMENT   Other_GSSs: BOGRE19TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            Seq primer: TR
            Class: sheared ends.
            DNA is from a doubled haploid provided by Tom Osborn.
FEATURES   Location/Qualifiers
     source   1..833
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGRE19"
                     /clone_lib="BOGR"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT   152 a   226 c   214 g   241 t
ORIGIN

RESULT 14

Qy   3 ATTCTAGTCGACAAGAGTTTGTCAGTGG 30
        | ||||| |||||||| ||| ||||||||
Db   8 AATCTAGCCGTCAAGTGTCGGTCAGTGG 35

Query Match   60.6%; Score 20; DB 17; Length 833;
Best Local Similarity   82.1%; Pred. No. 3.6e+02;
Matches 23; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

LOCUS   BM468802   978 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION AGENCOURT_6445892 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5587137
            5', mRNA sequence.
ACCESSION   BM468802
VERSION   BM468802.1   GI:18517844
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 978)
  AUTHORS   NIH-MGC http://mgc.ncl.nlh.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) .
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12356 row: c column: 10
            High quality sequence start: 92
            High quality sequence stop: 260.
FEATURES   Location/Qualifiers
     source   1..978
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5587137"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC library."
BASE COUNT   283 a   224 c   240 g   231 t

RESULT 15

```
Qy         14 CAAGAGTTTGTCAGTGGGAG 33
              |||||||||||||||||||||
Db        785 CAAGAGTTTGTCAGTGGGAG 766
```

ORIGIN

Query Match   60.6%; Score 20; DB 13; Length 978;
Best Local Similarity   100.0%; Pred. No. 3.9e+02;
Matches 20; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

BB150049/c

LOCUS   BB150049   240 bp   mRNA   linear   EST 28-JUN-2000
DEFINITION   BB150049 RIKEN full-length enriched, 6 days neonate skin Mus
             musculus cDNA clone A030003L20 3', mRNA sequence.
ACCESSION   BB150049
VERSION   BB150049.1 GI:8804986
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 240)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
             ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
             Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
             Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
             Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe.M.
             , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura.M., Oda,H., Okazaki,Y.
             , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
             ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
             , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
             ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
             , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
             ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp.
             URL:http://genome.gsc.riken.go.jp/
             Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
             ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Thermostabilization and thermoactivation of thermolabile enzymes by
             trehalose and its application for the synthesis of full length
             cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
             Tomaru,Y., Carninci,P., Shibata,K., Ozawa,Y., Muramatsu,M., Okazaki
             ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
             system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
             19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
             further details.
FEATURES   Location/Qualifiers
     source   1..240
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A030003L20"
                     /clone_lib="RIKEN full-length enriched, 6 days neonate
                     skin"
                     /tissue_type="skin"
                     /dev_stage="6 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
```

prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGGAGGATCCAAGAGCTGCTTTTTTTTTTTTTTTTVN 3'), cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGGATTCTCGAGTTAATTAAATTAAATCCCCCCCCCCC
3'). cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT      65 a    43 c    47 g    85 t
ORIGIN

Query Match      60.0%;  Score 19.8; DB 10; Length 240;
Best Local Similarity    77.4%;  Pred. No. 2.2e+02;
Matches  24;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

```
Qy          1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGG 31
              ||||||||  | |||| | |||||||||||||
Db         43 GCATTCTAGGGGCCAAGGATGTCTCAGTGGG 13
```

Search completed: December 16, 2002, 11:43:09
Job time : 2433 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: December 16, 2002, 10:54:58 ; Search time 32 Seconds
                      (without alignments)
                      407.819 Million cell updates/sec

Title: US-09-435-471B-9
Perfect score: 33
Sequence: 1 gcattctagtcgacaagagtttgtcagtgggag 33
Scoring table: IDENTITY_NUC
                      Gapop 10.0 , Gapext 1.0
Searched: 355320 seqs, 197730502 residues
Total number of hits satisfying chosen parameters: 710640
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                      Maximum Match 100%
                      Listing first 45 summaries

Database :
Published_Applications_NA:*
1: /cgn2_6/ptcdata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptcdata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptcdata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptcdata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptcdata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptcdata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptcdata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptcdata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptcdata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptcdata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptcdata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptcdata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptcdata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptcdata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19.8	60.0	401	9	US-09-946-807-666	Sequence 666, Appli
2	19.8	60.0	401	10	US-09-946-807-667	Sequence 667, App
3	19.8	60.0	401	10	US-09-795-668-666	Sequence 666, App
4	19.8	60.0	401	10	US-09-795-668-667	Sequence 667, App
5	19.8	60.0	401	10	US-09-795-686-666	Sequence 666, App
6	19.8	60.0	401	10	US-09-795-686-667	Sequence 667, App
7	19.8	60.0	401	10	US-09-946-807-1	Sequence 1, Appli
8	19.8	60.0	1503841	10	US-09-946-807-1	Sequence 1, Appli
9	19.8	60.0	1503841	10	US-09-795-668-1	Sequence 1, Appli
10	19.8	60.0	1503841	10	US-09-795-686-1	Sequence 1, Appli
11	18.6	56.4	709	10	US-09-960-253-137	Sequence 137, App
12	18.6	56.4	1732	9	US-10-063-547-81	Sequence 81, Appl
13	18.6	56.4	1732	12	US-10-006-867-81	Sequence 81, Appl
14	18.6	56.4	1732	12	US-10-052-586-281	Sequence 281, Appl
15	18.6	56.4	3073	12	US-09-867-701-10910	Sequence 10910, A
16	18.6	56.4	3073	10	US-09-920-300A-1690	Sequence 1690, Ap
17	18.4	55.8	3073	12	US-10-033-528-1690	Sequence 1690, Ap
18	18.4	55.8	1311	10	US-09-964-824A-232	Sequence 232, App
19	18.4	55.8	2484	10	US-09-839-497A-4	Sequence 4, Appli
20	18.4	55.8	148567	10	US-09-801-876B-3	Sequence 3, Appli
21	18.2	55.2	43058	10	US-09-954-456-292	Sequence 292, App
22	18.2	55.2	43058	10	US-09-954-456-529	Sequence 529, App
23	18.2	55.2	43058	10	US-09-880-107-3950	Sequence 3950, Ap
24	18.2	55.2	659158	9	US-09-771-208-20	Sequence 20, Appl
25	18	54.5	368	10	US-10-960-352-2011	Sequence 2011, Ap
26	18	54.5	1786	12	US-10-052-586-481	Sequence 481, App
27	17.6	53.3	1803	10	US-09-801-368-157	Sequence 157, App
28	17.6	53.3	1858	9	US-09-965-529-58	Sequence 58, Appl
29	17.6	53.3	2000	9	US-09-938-842A-3773	Sequence 3773, Ap
30	17.6	53.3	152331	9	US-10-095-407-16	Sequence 16, Appl
31	17.6	53.3	176373	9	US-10-095-407-17	Sequence 17, Appl
32	17.4	52.7	196	10	US-09-878-574-9563	Sequence 9563, Ap
33	17.4	52.7	404	12	US-09-728-445-1182	Sequence 1182, App
34	17.4	52.7	5866	12	US-10-044-090-215	Sequence 215, App
35	17.2	52.1	144	10	US-09-864-761-18243	Sequence 18243, A
36	17.2	52.1	382	10	US-09-770-791-163	Sequence 163, App
37	17.2	52.1	387	10	US-09-864-761-16334	Sequence 16334, A
38	17.2	52.1	470	10	US-09-864-761-1484	Sequence 1484, Ap
39	17.2	52.1	527	10	US-09-864-761-32816	Sequence 32816, A
40	17.2	52.1	657	9	US-09-938-842A-3874	Sequence 3874, Ap
41	17.2	52.1	2895	9	US-09-938-842A-2540	Sequence 2540, A
42	17.2	52.1	6319	10	US-09-070-927A-17	Sequence 17, Appl
43	17.2	52.1	32768	10	US-10-095-494-58	Sequence 58, Appl
44	17	51.5	442	10	US-09-864-761-224	Sequence 224, App
45	17	51.5	442	10	US-09-864-761-225	Sequence 225, App

ALIGNMENTS

RESULT 1
US-09-946-807-667/c
Sequence 666, Application US/09946807
Patent No. US2002016514A1

GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinthorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
FILE REFERENCE: 2345.2004-001
CURRENT APPLICATION NUMBER: US/09/946,807
CURRENT FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: US/09/795,668
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,716
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 666
LENGTH: 401
TYPE: DNA
ORGANISM: Homo sapiens
US-09-946-807-666

Query Match 60.0%; Score 19.8; DB 9; Length 401;
Best Local Similarity 77.4%;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy    2 CATTCTAGTCGACACAGAGTTTGTCAGTGGGA 32
         || ||||| ||| | | ||||||||||||
Db  167 CCTTTTAGTCTTCAACAAATTGTCAGTGGGA 137
```

RESULT 2
US-09-946-807-667/c
Sequence 667, Application US/09946807
Patent No. US2002016514A1

GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinthorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 667
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-667

```
Query Match             60.0%;  Score 19.8;  DB 9;  Length 401;
Best Local Similarity   77.4%;  Pred. No. 5.2;
Matches   24;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    2 CATTCTAGTCGACAAGAGTTTGTTCAGTGGGA 32
          || |||||  ||| |  ||||||||||||||||
Db   52 CCTTTTAGTCTTTCAACAAATTGTCAGTGGGA 22
```

RESULT 3
US-09-795-668-666/c
Sequence 666, Application US/09795668
; Patent No. US20020045577A1
GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 666
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-666

```
Query Match             60.0%;  Score 19.8;  DB 10;  Length 401;
Best Local Similarity   77.4%;  Pred. No. 5.2;
Matches   24;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    2 CATTCTAGTCGACAAGAGTTTGTTCAGTGGGA 32
          || |||||  ||| |  ||||||||||||||||
Db  167 CCTTTTAGTCTTTCAACAAATTGTCAGTGGGA 137
```

RESULT 4
US-09-795-668-667/c
Sequence 667, Application US/09795668
; Patent No. US20020045577A1
GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001

RESULT 5
US-09-795-686-666/c
Sequence 666, Application US/09795686
; Patent No. US20020094954A1
GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 666
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-666

```
Query Match             60.0%;  Score 19.8;  DB 10;  Length 401;
Best Local Similarity   77.4%;  Pred. No. 5.2;
Matches   24;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    2 CATTCTAGTCGACAAGAGTTTGTTCAGTGGGA 32
          || |||||  ||| |  ||||||||||||||||
Db   52 CCTTTTAGTCTTTCAACAAATTGTCAGTGGGA 22
```

RESULT 6
US-09-795-686-667/c
Sequence 667, Application US/09795686
; Patent No. US20020094954A1
GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 667
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-667

```
Query Match             60.0%;  Score 19.8;  DB 10;  Length 401;
Best Local Similarity   77.4%;  Pred. No. 5.2;
Matches   24;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    2 CATTCTAGTCGACAAGAGTTTGTTCAGTGGGA 32
          || |||||  ||| |  ||||||||||||||||
Db  167 CCTTTTAGTCTTTCAACAAATTGTCAGTGGGA 137
```

RESULT 7
US-09-946-807-1/c
Sequence 1, Application US/09946807
Patent No. US2002016514A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinthorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
FILE REFERENCE: 2345.2004-001
CURRENT APPLICATION NUMBER: US/09/946,807
CURRENT FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: US/09/795,668
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,716
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1503841
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: r=g or a
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: y=t/u or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: k=g or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: m=a or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: s=g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: w=a or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: b=g or c or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: d=a or g or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: h=a or c or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: v=a or g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1

Query Match 60.0%; Score 19.8; DB 9; Length 1503841;
Best Local Similarity 77.4%; Pred. No. 27;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
        | || ||||| |||| | || |||||||||||||
Db 815361 CCTTTTAGTCTTCAACAAATTGTCAGTGGGA 815331

RESULT 8
US-09-795-668-1/c
Sequence 1, Application US/09795668
Patent No. US2002004577A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinthorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
FILE REFERENCE: 2345.2004-001
CURRENT APPLICATION NUMBER: US/09/795,668
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,716
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1503841
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: r=g or a
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: y=t/u or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: k=g or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: m=a or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: s=g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: w=a or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: b=g or c or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: d=a or g or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: h=a or c or t/u
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: v=a or g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(1531)
OTHER INFORMATION: n=a or g or c or t/u
US-09-795-668-1

Query Match 60.0%; Score 19.8; DB 10; Length 1503841;
Best Local Similarity 77.4%; Pred. No. 27;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy   2 CATTCTAGTCGACACGAGTTTGTCAGTGGGA 32
        | ||| ||||| ||| | |||||||||||||
Db 815361 CCTTTTAGTCTTCAACAAATTGTCAGTGGGA 815331

RESULT 9
US-09-795-686-1/c

Sequence 1, Application US/09795686
Patent NO. US200200094954A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn
APPLICANT: Steinthorsdottir, Valgerdur
APPLICANT: Gulcher, Jeffrey R.
FILE REFERENCE: 2345.2005-001
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
CURRENT APPLICATION NUMBER: US/09/795,686
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 09/515,715
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 1531
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
  LENGTH: 1503841
  TYPE: DNA
  ORGANISM: Homo sapiens
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: r=g or a
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: y=t/u or c
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: m=a or c
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: k=g or t/u
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: s=g or c
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: w=a or t/u
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: b=g or c or t/u
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: d=a or g or t/u
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: h=a or c or t/u
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: v=a or g or c
  NAME/KEY: misc_feature
  LOCATION: (1)..(1531)
  OTHER INFORMATION: n=a or g or c or t/u

US-09-795-686-1

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
                    |||||| ||||| ||| | |||||||||||||
Db 815361 CCTTTTAGTCTTCAACAAATTGTCAGTGGGA 815331

Query Match   60.0%;   Score 19.8;   DB 10;   Length 1503841;
Best Local Similarity   77.4%;   Pred. No. 27;
  Matches   24;   Conservative   0;   Mismatches   7;   Indels   0;   Gaps   0;

RESULT 10
US-09-960-253-137
Sequence 137, Application US/09960253
Patent NO. US2002012361A9A1
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Mohamath, Raodoh
APPLICANT: Ledes, Michael J.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.556
CURRENT APPLICATION NUMBER: US/09/960,253
CURRENT FILING DATE: 2001-09-20
NUMBER OF SEQ ID NOS: 187
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 137
  LENGTH: 709
  TYPE: DNA
  ORGANISM: Homo saplens

US-09-960-253-137

Qy   4 TTCTAGTCGACAAGAGTTTGTCAGT 28
             ||||||| || | |||||||||||||
Db 606 TTCTAGGCGTGAGGAGTTTGTCAGT 630

Query Match   56.4%;   Score 18.6;   DB 10;   Length 709;
Best Local Similarity   84.0%;   Pred. No. 21;
  Matches   21;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

RESULT 11
US-10-063-547-81/c
Sequence 81, Application US/10063547
Publication NO. US20020182638A1
GENERAL INFORMATION:
APPLICANT: Eaton,Dan L.
APPLICANT: Filvaroff,Ellen
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Grimaldi,Christopher J.
APPLICANT: Gurney,Austin L.
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,547
CURRENT FILING DATE: 2002-05-02
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 81
  LENGTH: 1732
  TYPE: DNA
  ORGANISM: Homo Sapien

US-10-063-547-81

Qy   1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGGAG 33
             |||||||| || | |||||||||||||||||||
Db   808 GCAGGCTTGGCAGCAGGAGTCTGGCAGTGGGGAG 776

Query Match   56.4%;   Score 18.6;   DB 9;   Length 1732;
Best Local Similarity   72.7%;   Pred. No. 25;
  Matches   24;   Conservative   0;   Mismatches   9;   Indels   0;   Gaps   0;

RESULT 12
US-10-006-867-81/c
Sequence 81, Application US/10006867
Patent NO. US20020119130A1
GENERAL INFORMATION:
APPLICANT: Eaton,Dan L.
APPLICANT: Filvaroff,Ellen
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Grimaldi,Christopher J.
APPLICANT: Gurney,Austin L.
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3230R1C1

CURRENT APPLICATION NUMBER: US/10/006,867
CURRENT FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: 60/063435
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/064215
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088740
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088811
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088825
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088863
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090688
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091628
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/096012
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096757
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/096949
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/096959
PRIOR FILING DATE: 1998-08-18
PRIOR APPLICATION NUMBER: 60/097954
PRIOR FILING DATE: 1998-08-18
PRIOR APPLICATION NUMBER: 60/097971
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/097979
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098741
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/099741
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/099763
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099792
PRIOR FILING DATE: 1998-09-10

PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099815
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100662
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100683
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100684
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/100930
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101279
PRIOR FILING DATE: 1998-09-22
PRIOR APPLICATION NUMBER: 60/101475
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101743
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/102570
PRIOR FILING DATE: 1998-09-30
PRIOR APPLICATION NUMBER: 60/103449
PRIOR FILING DATE: 1998-10-06
PRIOR APPLICATION NUMBER: 60/103678
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103679
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/103711
PRIOR FILING DATE: 1998-10-08
PRIOR APPLICATION NUMBER: 60/105000
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105002
PRIOR FILING DATE: 1998-10-20
PRIOR APPLICATION NUMBER: 60/105881
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/106030
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106464
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 60/106856
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/108807
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/112419
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/112422
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/112854
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/112853
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113011
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113300
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113408
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113430
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/114223
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115614
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 60/116527
PRIOR FILING DATE: 1999-01-12
PRIOR FILING DATE: 1999-01-20

RESULT 13
US-10-052-586-281/c
Sequence 281, Application US20020127584A1
Patent No. US20020127584A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10,052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266

Query Match   56.4%; Score 18.6; DB 12; Length 1732;
Best Local Similarity   72.7%; Pred. No. 9;
Matches   24; Conservative   0; Mismatches   25; Indels   0; Gaps   0;

Qy   1 GCATTCTAGTCGACACAGAGTTTGTCAGTGGGAG   33
          ||| || |||||| |||| |||||| |||||||||
Db   808 GCAGGCTTGGCACCAGGAGTCTGGCAGTGGGAG   776

PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21

PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/119285
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119287
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119525
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/115549
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/120014
PRIOR FILING DATE: 1999-02-11
PRIOR APPLICATION NUMBER: 60/120122
PRIOR FILING DATE: 1999-02-11
PRIOR APPLICATION NUMBER: 60/129122
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 60/129674
PRIOR FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/138387
PRIOR FILING DATE: 1999-06-09
PRIOR APPLICATION NUMBER: 60/144791
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/169495
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 60/175481
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: 60/191007
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 60/199397
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/193397
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/380139
PRIOR FILING DATE: 1998-08-25
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380142

PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088722
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10

PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088740
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088811
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088825
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088863
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089090
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089908

Qy   1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGGAG 33
           || |||| || | | || | |||| |||||||
Db   808 GCAGGCTTGGCGAGGAGGAGTCTGGCAGTGGGAG 776

Query Match   56.4%;   Score 18.6;   DB 12;   Length 1732;
Best Local Similarity 72.7%;   Pred. No. 25;
Matches   24;   Conservative   0;   Mismatches   9;   Indels   0;   Gaps   0;

RESULT 14
US-09-867-701-10910
Sequence 10910, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
 APPLICANT: Aglate, Paul A.
 APPLICANT: Jones, Robert
 APPLICANT: Harlocker, Susan L.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
 FILE REFERENCE: 210121.497
 CURRENT APPLICATION NUMBER: US/09/867,701
 CURRENT FILING DATE: 2001-05-29
 NUMBER OF SEQ ID NOS: 10912
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEQ ID NO 10910
 LENGTH: 3073
 TYPE: DNA
 ORGANISM: Homo sapiens
US-09-867-701-10910

Qy   4 TTCTAGTCGACAAGAGTTTGTCAGT 28
           |||||| || |||||||||||||
Db   606 TTCTAGCGCGTGAGGAGTTTGTCAGT 630

Query Match   56.4%;   Score 18.6;   DB 10;   Length 3073;
Best Local Similarity 84.0%;   Pred. No. 28;
Matches   21;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

RESULT 15

US-09-920-300A-1690
; Sequence 1690, Application US/09920300A
; Patent No. US20020136728A1
; GENERAL INFORMATION:
; APPLICANT: King, Gordon E.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Xu, Jiangchun
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.547
; CURRENT APPLICATION NUMBER: US/09/920,300A
; CURRENT FILING DATE: 2001-07-31
; NUMBER OF SEQ ID NOS: 1789
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1690
; LENGTH: 3073
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-300A-1690

Query Match   56.4%;   Score 18.6;   DB 10;   Length 3073;
Best Local Similarity   84.0%;   Pred. No. 28;
Matches   21;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

Qy   4 TTCTAGTCGACAAAGAGTTTGTCAGT 28
              ||||| |||| || |||||||||||||
Db   606 TTCTAGGCCTGAGGAGTTGTCAGT 630

Search completed: December 16, 2002, 11:13:13
Job time : 998 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3

Run on: December 16, 2002, 10:54:58 ; Search time 302 Seconds

Sequence: 1 gcattctagtcgacaagagtttgtcagtgggag 33

Perfect score: 33

Title: US-09-435-471b-9

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101102:*
```
 1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
 2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
 3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
 4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
 5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
 6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
 7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
 8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
 9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	19.8	60.0	401	22	AAK95867	Human neuregulin g
2	19.8	60.0	401	22	AAK95868	Human neuregulin g
3	19.8	60.0	401	22	AAK97360	Human neuregulin g
4	19.8	60.0	401	22	AAK97361	Human neuregulin g
5	19.8	60.0	1715	23	ABL29013	Drosophila melanog
6	19.8	60.0	3792	23	ABL29012	Drosophila melanog
7	19.8	60.0	1503900	22	AAK95240	Human neuregulin-1
8	19.8	60.0	1503900	22	AAK96733	Human neuregulin-1
9	19.2	58.2	731	21	AAC43835	Arabidopsis thalia
10	19.2	58.2	16236	24	AAS74037	DNA encoding novel
11	19.2	58.2	782	23	ABL13023	Human immune syste
12	19	57.6	1177	9	AAN81277	New sequence of ra
13	19	57.6	1177	17	AAT18809	Type I (beta-1) ra
14	19	57.6	1802	9	AAN82021	Sequence encoding
15	19	57.6	2600	10	AAN90750	Protein kinase C b
16	19	57.6	3202	17	AAQ74001	Type I (beta-1) ra
17	19	57.6	3418	17	AAN81279	Entire nucleotide
18	19	57.6	3418	17	AAT15006	Type II (beta-2) r
19	19	57.6	4829	20	AAZ35666	Ovine interferon g
20	19	57.6	4842	20	AAZ35665	Ovine interferon g
21	19	57.6	925	21	AAC37620	Arabidopsis thalia
22	18.8	57.0	1331	21	AAC39790	Arabidopsis thalia
23	18.8	57.0	1465	21	AAC51439	Arabidopsis thalia
24	18.8	57.0	6081	24	ABN80193	Human chemically m
25	18.8	57.0	6081	24	ABL34429	Human immune syste
26	18.6	56.4	666	24	AAQ60322	Human colon cancer
27	18.6	56.4	709	24	AAK70266	Human lung cancer
28	18.6	56.4	824	22	AAI93936	Human neuroblastom
29	18.6	56.4	1720	22	AAC84105	Human EXCS encodin
30	18.6	56.4	1731	22	AAS46065	Human DNA encoding
31	18.6	56.4	1732	21	AAA37064	Human PRO1557 (UNQ
32	18.6	56.4	1732	22	AAF92098	Human PRO1557 cDNA
33	18.6	56.4	1738	24	AAF54300	DNA encoding prote
34	18.6	56.4	1738	24	ABK12126	Human cDNA encodin
35	18.6	56.4	2639	23	ABL02633	Drosophila melanog
36	18.6	56.4	3073	24	ABK46139	cDNA encoding colo
37	18.6	56.4	3073	24	ABL87932	cDNA encoding ovar
38	18.6	56.4	4945	23	ABL02632	Drosophila melanog
39	18.6	56.4	5669	22	AAK70832	Human immune/haema
40	18.6	56.4	17450	22	AAL05953	Human reproductive
41	18.6	56.4	17450	22	AAK70834	Human immune/haema
42	18.4	56.4	17450	23	ABL98517	Human testicular a
43	18.4	56.0	60	24	ABN36300	Human spliced tran
44	18.4	55.8	445	20	AAX51871	Human secreted pro
45	18.4	55.8	460	20	AAX51869	Human secreted pro

ALIGNMENTS

```
RESULT 1
AAK95867/c
ID   AAK95867 standard; DNA; 401 BP.
XX
AC   AAK95867;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin gene single nucleotide polymorphism SNP8NRG815395.
XX
KW   Human; neuregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene;
KW   gene therapy; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164876-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06376.
XX
PR   28-FEB-2000; 2000US-0515715.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR   WPI; 2001-550179/61.
XX
PT   Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT   preventing diagnosing and treating schizophrenia -
XX
```

RESULT 2

AAK95868/c

ID   AAK95868 standard; DNA; 401 BP.
XX
AC   AAK95868;
XX
DT   17-DEC-2001 (first entry)
XX
DE   Human neuregulin gene single nucleotide polymorphism SNP8NRG815510.
XX
KW   Human; neuregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene;
KW   gene therapy; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164876-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06376.
XX
PR   28-FEB-2000; 2000US-0515715.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR   WPI; 2001-550179/61.
XX
PT   Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT   preventing diagnosing and treating schizophrenia -
XX
PS   Disclosure; Page 609; 750pp; English.
XX
CC   This sequence represents a single nucleotide polymorphism (SNP) of the
CC   human neuregulin-1 associated gene 1 (NRG1AG1) of the invention. The
CC   NRG1AG1 gene is also referred to as the human Schizophrenia gene. The
CC   invention also relates to fragments or variants of the gene and the
CC   NRG1AG1 polypeptides they encode. The NRG1AG1 nucleic acids and
CC   polypeptides may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate NRG1AG1 expression. For example,
CC   they may be used to treat disorders associated with decreased expression
CC   by rectifying mutations or deletions in a patient's genome that affect
CC   the activity of NRG1AG1 by expressing inactive proteins or to supplement
CC   the patients own production of NRG1AG1. Additionally, the gene may be
CC   used to produce NRG1AG1 polypeptides, by inserting the nucleic acids into
CC   a host cell and culturing the cell to express the protein. The gene may
CC   also be used as DNA probes and primers in diagnostic assays to detect and
CC   quantitate the presence of similar nucleic acids in samples, and
CC   therefore which patients may be in need of restorative therapy. The
CC   NRG1AG1 polypeptides may also be used as antigens in the production of
CC   antibodies against NRG1AG1 and in assays to identify modulators of
CC   NRG1AG1 expression and activity. Anti-NRG1AG1 antibodies and antagonists
CC   may also be used to down regulate expression and activity. Anti-NRG1AG1
CC   antibodies may also be used as diagnostic agents for detecting the
CC   presence of NRG1AG1 polypeptides in samples. NRG1AG1 is associated with
CC   schizophrenia which may be prevented, diagnosed and/or treated by the
CC   above methods.
XX
SQ   Sequence 401 BP; 138 A; 91 C; 67 G; 103 T; 2 other;

Query Match   60.0%; Score 19.8; DB 22; Length 401;
Best Local Similarity   77.4%;
Matches 24; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
              ||||| ||| || || | | ||||||||||
Db   167 CCTTTTAGTCTTCAACAAATTGTCAGTGGGA 137

RESULT 3

AAK97360/c

ID   AAK97360 standard; DNA; 401 BP.
XX
AC   AAK97360;
XX
DT   17-DEC-2001 (first entry)
XX
DE   Human neuregulin gene single nucleotide polymorphism SNP8NRG815395.
XX
KW   Human; neuregulin 1 gene; schizophrenia; gene therapy; SNP;
KW   single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164877-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06377.
XX
PR   28-FEB-2000; 2000US-0515716.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR   WPI; 2001-51481/56.
XX
PT   Neuregulin 1 nucleic acids and proteins useful for diagnosing
PT   preventing and treating schizophrenia -
XX
PS   Disclosure; Page 194; 756pp; English.
XX
CC   This sequence represents a single nucleotide polymorphism (SNP)
CC   from the human neuregulin 1 gene of the invention.
CC   The invention also relates to fragments or variants of the neuregulin 1
CC   gene. The gene and its proteins may be used in the prevention, diagnosis
CC   human neuregulin-1 associated gene 1 (NRG1AG1) of the invention. The
CC   NRG1AG1 gene is also referred to as the human Schizophrenia gene. The
CC   invention also relates to fragments or variants of the gene and the
CC   NRG1AG1 polypeptides they encode. The NRG1AG1 nucleic acids and
CC   polypeptides may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate NRG1AG1 expression. For example,
CC   they may be used to treat disorders associated with decreased expression
CC   by rectifying mutations or deletions in a patient's genome that affect
CC   the activity of NRG1AG1 by expressing inactive proteins or to supplement
CC   the patients own production of NRG1AG1. Additionally, the gene may be
CC   used to produce NRG1AG1 polypeptides, by inserting the nucleic acids into
CC   a host cell and culturing the cell to express the protein. The gene may
CC   also be used as DNA probes and primers in diagnostic assays to detect and
CC   quantitate the presence of similar nucleic acids in samples, and
CC   therefore which patients may be in need of restorative therapy. The
CC   NRG1AG1 polypeptides may also be used as antigens in the production of
CC   antibodies against NRG1AG1 and in assays to identify modulators of
CC   NRG1AG1 expression and activity. Anti-NRG1AG1 antibodies and antagonists
CC   may also be used to down regulate expression and activity. Anti-NRG1AG1
CC   antibodies may also be used as diagnostic agents for detecting the
CC   presence of NRG1AG1 polypeptides in samples. NRG1AG1 is associated with
CC   schizophrenia which may be prevented, diagnosed and/or treated by the
CC   above methods.
XX
SQ   Sequence 401 BP; 133 A; 99 C; 71 G; 95 T; 3 other;

Query Match   60.0%; Score 19.8; DB 22; Length 401;
Best Local Similarity   77.4%;
Matches 24; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
              ||||| ||| || || | | ||||||||||
Db   52 CCTTTTAGTCTTCAACAAATTGTCAGTGGGA 22

```
CC   and treatment of diseases associated with inappropriate neuregulin 1
CC   expression, such as schizophrenia. For example they may be used to treat
CC   disorders associated with decreased neuregulin 1 expression by rectifying
CC   mutations or deletions in a patient's genome that affect the activity of
CC   neuregulin 1 by expressing inactive proteins or to supplement the
CC   patients own production of polypeptides. Additionally, the gene may be
CC   used to produce the neuregulin 1 protein, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. The gene
CC   and its complementary sequences may also be used as DNA probes in
CC   diagnostic assays to detect and quantitate the presence of similar
CC   nucleic acids in samples, and therefore which patients may be in need of
CC   restorative therapy. The protein may also be used as antigens in the
CC   production of antibodies against neuregulin 1 and in assays to identify
CC   modulators of neuregulin 1 expression and activity. The antibodies and
CC   antagonists may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of neuregulin 1 in samples.
XX
SQ   Sequence 401 BP; 138 A; 91 C; 67 G; 103 T; 2 other;

Qy        2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
            |||||  |||| ||| |||||  |||||||||||
Db      167 CCTTTTAGTCTTCAACAAATTGTCAGTGCGGA 137

Query Match             60.0%;  Score 19.8;  DB 22;  Length 401;
Best Local Similarity   77.4%;  Pred. No. 18;
Matches  24;  Conservative   0;  Mismatches    7;  Indels    0;  Gaps    0;


RESULT 4
AAK97361/c
ID   AAK97361 standard; DNA; 401 BP.
XX
AC   AAK97361;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin gene single nucleotide polymorphism SNP8NRG815510.
XX
KW   Human; neuregulin 1 gene; schizophrenia; gene therapy; SNP;
KW   single nucleotide polymorphism; ds;
XX
OS   Homo sapiens.
XX
PN   WO200164877-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001;  2001WO-US06377.
XX
PR   28-FEB-2000;  2000US-0515716.
XX
PA   (DECO-)  DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR   WPI; 2001-51484l/56.
XX
PT   Neuregulin 1 nucleic acids and proteins useful for diagnosing
PT   preventing and treating schizophrenia -
XX
PS   Disclosure; Page 194; 756pp; English.
XX
CC   This sequence represents a single nucleotide polymorphism (SNP)
CC   from the human neuregulin 1 gene of the invention.
CC   The invention also relates to fragments or variants of the invention.
CC   The gene and its proteins may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate neuregulin 1
CC   expression, such as schizophrenia. For example they may be used to treat
CC   disorders associated with decreased neuregulin 1 expression by rectifying
CC   mutations or deletions in a patient's genome that affect the activity of
CC   neuregulin 1 by expressing inactive proteins or to supplement the
CC   patients own production of polypeptides. Additionally, the gene may be
```

```
CC   used to produce the neuregulin 1 protein, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. The gene
CC   and its complementary sequences may also be used as DNA probes in
CC   diagnostic assays to detect and quantitate the presence of similar
CC   nucleic acids in samples, and therefore which patients may be in need of
CC   restorative therapy. The protein may also be used as antigens in the
CC   production of antibodies against neuregulin 1 and in assays to identify
CC   modulators of neuregulin 1 expression and activity. The antibodies and
CC   antagonists may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of neuregulin 1 in samples.
XX
SQ   Sequence 401 BP; 133 A; 99 C; 71 G; 95 T; 3 other;

Qy        2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
            |||||  |||| ||| |||||  |||||||||||
Db       52 CCTTTTAGTCTTCAACAAATTGTCAGTGCGGA 22

Query Match             60.0%;  Score 19.8;  DB 22;  Length 401;
Best Local Similarity   77.4%;  Pred. No. 18;
Matches  24;  Conservative   0;  Mismatches    7;  Indels    0;  Gaps    0;


RESULT 5
ABL29013/c
ID   ABL29013 standard; DNA; 1715 BP.
XX
AC   ABL29013;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 38512.
XX
KW   Drosophila; developmental biology; cell signalling; Insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001;  2001WO-US09231.
XX
PR   23-MAR-2000;  2000US-191639P.
PR   11-JUL-2000;  2000US-0614150.
XX
PA   (PEKE )  PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 38512; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1715 BP; 436 A; 426 C; 460 G; 393 T; 0 other;
```

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
            ||||||||||||||||||||||||||||| |
Db   842 CATTATGGTGGACAAGAGATTTCGTCGGTGGCA 812

RESULT 6

ID   ABL29012 standard; DNA; 3792 BP.
XX
AC   ABL29012;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 38509.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-1916937P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PS   Claim 1; SEQ ID NO 38509; 21pp + Sequence Listing; English.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3792 BP; 1016 A; 905 C; 846 G; 1025 T; 0 other;

Query Match   60.0%; Score 19.8; DB 23; Length 3792;
Best Local Similarity   77.4%; Pred. No. 24;
Matches   24; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
            |||||| ||||||||||||||||| |||| |
Db   1951 CATTATGGTGGACAAGAGATTTCGTCGGTGGCA 1981

RESULT 7

ID   AAK95240 standard; DNA; 1503900 BP.
XX
AC   AAK95240;

XX
DT   17-DEC-2001 (first entry)
XX
DE   Human neuregulin-1 gene.
XX
KW   Human; neuregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene;
KW   gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO20016876-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06376.
XX
PR   28-FEB-2000; 2000US-0515715.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
PT   Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT   preventing diagnosing and treating schizophrenia -
XX
PS   Disclosure; Page 90-501; 750pp; English.
XX
CC   This sequence represents the human neuregulin-1 associated gene 1
CC   (NRG1AG1) of the invention. The NRG1AG1 gene is also referred to as the
CC   human Schizophrenia gene. The invention also relates to fragments or
CC   variants of the gene and the NRG1AG1 polypeptides they encode. The
CC   NRG1AG1 nucleic acids and polypeptides may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate NRG1AG1
CC   expression. For example, they may be used to treat disorders associated
CC   with decreased expression by rectifying mutations or deletions in a
CC   patient's genome that affect the activity of NRG1AG1 by expressing
CC   inactive proteins or to supplement the patients own production of
CC   NRG1AG1. Additionally, the gene may be used to produce NRG1AG1
CC   polypeptides, by inserting the nucleic acids into a host cell and
CC   culturing the cell to express the protein. The gene may also be used as
CC   DNA probes and primers in diagnostic assays to detect and quantitate the
CC   presence of similar nucleic acids in samples, and therefore which
CC   patients may be in need of restorative therapy. The NRG1AG1 polypeptides
CC   may also be used as antigens in the production of antibodies against
CC   NRG1AG1 and in assays to identify modulators of NRG1AG1 expression and
CC   activity. Anti-NRG1AG1 antibodies and antagonists may also be used to
CC   down regulate expression and activity. Anti-NRG1AG1 antibodies may
CC   also be used as diagnostic agents for detecting the presence of NRG1AG1
CC   polypeptides in samples. NRG1AG1 is associated with schizophrenia which
CC   may be prevented, diagnosed and/or treated by the above methods.
XX
WPI   2001-550179/61.
DR   P-PSDB; AAG67900, AAG67901, AAG67902, AAG67903, AAG67904, AAG67905,
DR   AAG67906, AAG67907, AAG67908, AAG67909, AAG67910, AAG67911, AAG67912,
DR   AAG67913, AAG67914, AAG67915, AAG67916, AAG67917, AAG67918, AAG67919,
DR   AAG67920, AAG67921, AAG67922, AAG67923, AAG67924, AAG67925, AAG67926,
DR   AAG67927, AAG67928, AAG67929, AAG67930, AAG67931, AAG67932, AAG67933,
DR   AAG67934, AAG67935, AAG67936, AAG67937.
XX
SQ   Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;

Query Match   60.0%; Score 19.8; DB 22; Length 1503900;
Best Local Similarity   77.4%; Pred. No. 99;
Matches   24; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
            ||||| ||| ||||||||||||||||||||
Db   815360 CCTTTTAGTCTTCAACAAAATGTCAGTGGGA 815330

RESULT 8

ID   AAK96733 standard; DNA; 1503900 BP.

XX
AC   AAK96733;
XX
DT   17-DEC-2001 (first entry)
XX
DE   Human neuregulin-1 gene.
XX
KW   Human; neuregulin 1 gene; schizophrenia; gene therapy; ds.
OS   Homo sapiens.
XX
PN   WO200164877-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06377.
XX
PR   28-FEB-2000; 2000US-0515716.
XX
DR   WPI; 2001-514841/56.
DR   P-PSDB; AAG67938, AAG67939, AAG67940, AAG67941, AAG67942, AAG67943,
DR   AAG67944, AAG67945, AAG67946, AAG67947, AAG67948, AAG67949,
DR   AAG67950, AAG67951, AAG67952, AAG67953, AAG67954, AAG67955,
DR   AAG67956, AAG67957, AAG67958, AAG67959, AAG67960, AAG67961,
DR   AAG67962, AAG67963, AAG67964, AAG67965, AAG67966, AAG67967,
DR   AAG67968, AAG67969, AAG67970, AAG67971, AAG67972, AAG67973,
DR   AAG67974, AAG67975.
XX
PI   Stefansson H, Steinthorsdottir V, Gulcher JR;
PA   (DECO-) DECODE GENETICS EHF.
XX
PT   Neuregulin 1 nucleic acids and proteins useful for diagnosing
PT   preventing and treating schizophrenia -
XX
PS   Disclosure; Page 345-756; 756pp; English.
XX
CC   This sequence represents the human neuregulin 1 gene of the invention.
CC   The invention also relates to fragments or variants of the neuregulin 1
CC   gene. The gene and its proteins may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate neuregulin 1
CC   expression, such as schizophrenia. For example they may be used to treat
CC   disorders associated with decreased neuregulin 1 expression by rectifying
CC   mutations or deletions in a patient's genome that affect the activity of
CC   neuregulin 1 by expressing inactive proteins or to supplement the
CC   patients own production of polypeptides. Additionally, the gene may be
CC   used to produce the neuregulin 1 protein, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. The gene
CC   and its complementary sequences may also be used as DNA probes in
CC   diagnostic assays to detect and quantitate the presence of similar
CC   nucleic acids in samples, and therefore which patients may be in need of
CC   restorative therapy. The protein may also be used as antigens in the
CC   production of antibodies against neuregulin 1 and in assays to identify
CC   modulators of neuregulin 1 expression and activity. The antibodies and
CC   antagonists may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of neuregulin 1 in samples.
XX
SQ   Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;

RESULT 9
AAC43835/c
ID   AAC43835 standard; DNA; 731 BP.
XX

Query Match   60.0%; Score 19.8; DB 22; Length 1503900;
Best Local Similarity   77.4%; Pred. No. 99;
Matches   24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
           ||||| ||||| |||| |||||||||||||||
Db 815360 CCTTTTAGTCGTTCAACAAATTGTCAGTGGGA 815330

AC   AAC43835;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 40686.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   07-MAY-1999; 99US-0132863.
PR   08-APR-1999; 99US-0128714.
PR   11-MAY-1999; 99US-0129845.
PR   14-MAY-1999; 99US-0130077.
PR   16-APR-1999; 99US-0134219.
PR   19-APR-1999; 99US-0134221.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0130991.
PR   30-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   04-MAY-1999; 99US-0132407.
PR   05-MAY-1999; 99US-0132484.
PR   06-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   07-MAY-1999; 99US-0132487.
PR   11-MAY-1999; 99US-0135629.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.

PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   20-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0145008.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145086.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145218.
PR   26-JUL-1999;   99US-0145224.
PR   27-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145518.
PR   27-JUL-1999;   99US-0145519.
PR   28-JUL-1999;   99US-0145913.
PR   28-JUL-1999;   99US-0145951.
PR   10-AUG-1999;   99US-0146171.
PR   11-AUG-1999;   99US-0148171.
PR   02-AUG-1999;   99US-0148819.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   03-AUG-1999;   99US-0146389.
PR   04-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   05-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   06-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   09-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   10-AUG-1999;   99US-0147935.
PR   13-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   16-AUG-1999;   99US-0148684.
PR   17-AUG-1999;   99US-0149368.
PR   18-AUG-1999;   99US-0149175.
PR   20-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   23-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   25-AUG-1999;   99US-0149930.
PR   .   99US-0150566.

PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151514.
PR   07-SEP-1999;   99US-0151930.
PR   10-SEP-1999;   99US-0153263.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0155596.
PR   29-SEP-1999;   99US-0157117.
PR   04-OCT-1999;   99US-0157753.
PR   05-OCT-1999;   99US-0157865.
PR   06-OCT-1999;   99US-0158029.
PR   07-OCT-1999;   99US-0158232.
PR   08-OCT-1999;   99US-0158369.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

Query Match   58.2%;   Score 19.2;   DB 21;   Length 731;
Best Local Similarity 87.5%;   Pred. No. 38;
Matches   21; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

```
Qy    3  ATTCTAGTCGACAAGAGTTTGTCA  26
            ||| ||||| ||||||||||| ||
Db  477  ATTGTAGTCAACAAGAGTTTGCCA  454
```

RESULT 10
AAS74037
ID   AAS74037 standard; cDNA; 782 BP.
XX
AC   AAS74037;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #9841.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG09850.
XX
PS Claim 1; SEQ ID No 984l; 103pp; English.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 782 BP; 177 A; 217 C; 243 G; 145 T; 0 other;

Query Match   58.2%; Score 19.2; DB 23; Length 782;
Best Local Similarity 75.0%; Pred. No. 39;
Matches   24; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

Qy   2 CATTCTAGTCGACAAAGAGAGTTTGTCAGTGGGAG 33
         |||| ||||| ||||| || ||| | |||||||||
Db 562 CCTTCCAGTCGCCAGGAATTCCTTAGTCGGGAG 593

RESULT 11
ABL33023
ID ABL33023 standard; DNA; 16236 BP.
AC ABL33023;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 996.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO20200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PS Claim 1; SEQ ID NO 996; 32pp + Sequence Listing; German.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 16236 BP; 4368 A; 355 C; 3510 G; 8003 T; 0 other;

Query Match   58.2%; Score 19.2; DB 24; Length 16236;
Best Local Similarity 75.0%; Pred. No. 74;
Matches   24; Conservative   0; Mismatches   8; Indels   0; Gaps   0;

Qy   1 GCATTCTAGTCGACAAAGAGAGTTTGTCAGTGGGA 32
         | ||| |||| ||| ||||||| || ||| |||||||
Db 6084 GGAATGTRAGTGGGATRAAGTGTTTGTGAGTTGGA 6115

RESULT 12
AAN81277/c
ID AAN81277 standard; cDNA; 1177 BP.
AC AAN81277;
XX
DT 24-OCT-1990 (first entry)
XX
DE New sequence of rat protein kinase C cDNA.
XX
KW Rat kinase protein C cDNA; enzyme; EC-2.7.1.37; pTB638; ss.
XX
OS Rat.
XX
FH Key   Location/Qualifiers
FT mat_peptide   1..675
FT   /*tag= a
FT polyA_signal   1156..1161
FT   /*tag= b
FT polyA_site   1177
FT   /*tag= c
XX
PN EP251244-A.
XX

RESULT 13
AAT18809/c
ID   AAT18809 standard; cDNA; 1177 BP.
XX
AC   AAT18809;
XX
DT   07-AUG-1996 (first entry)
XX
DE   Type I (beta-1) rat brain protein kinase C partial cDNA.
XX
KW   Protein kinase C; signal transduction; tumour; diagnosis; therapy;
XX
OS   Rattus sp.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..675
FT   /*tag= a
FT   polyA_signal   1156..1161
FT   /*tag= b
FT   misc_difference 1177
FT   /*tag= c
FT   /note= "base n at position 1177 signifies an
FT   unspecified number of adenine bases"
XX
PN   EP686695-A1.
XX
PD   13-DEC-1995.
XX
PF   26-JUN-1987; 87EP-0112109.
XX
PR   25-FEB-1987; 87JP-0040160.
PR   27-JUN-1986; 86JP-0149385.
PR   18-SEP-1986; 86JP-0217944.
PR   28-NOV-1986; 86JP-0281870.
XX
PA   (TAKE) TAKEDA CHEM IND LTD.
XX
PD   07-JAN-1988.
XX
PF   26-JUN-1987; 87EP-0109223.
XX
PR   25-FEB-1987; 87JP-0040160.
PR   27-JUN-1986; 86JP-0149385.
XX
XX
PA   (TAKE) TAKEDA CHEMICAL IND KK.
XX
PI   Ono Y, Kurokawa T, Igarashi K, Nishizuka Y;
XX
DR   WPI; 1988-001173/01.
DR   P-PSDB; AAP80695.
XX
PT   New human and rat protein kinase C -
PT   for studying cellular signal transductor mechanisms
XX
PS   Example; Fig 2.1 - 2.2; 32pp; English.
XX
CC   The SQ is from a rat brain mRNA-derived cDNA library. It is contained in
CC   plasmid pTB638. AA sequences of the peptides nos. 24 and 51 determined
CC   in Kikkawa [J. Biol. Chem 257, 1334 (1982)] stictly corresp. to the
CC   nucleotide sequences at nos. 445-480 and 220-312, respectively. Thus, the
CC   plasmid pTB638 was confirmed to be rat protein kinase C cDNA.
XX
SQ   Sequence 1177 BP; 319 A; 244 C; 277 G; 337 T; 0 other;
XX

Query Match   57.6%; Score 19; DB 9; Length 1177;
Best Local Similarity   100.0%; Pred. No. 53;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          15 AAGAGTTTGTCAGTGGGAG 33
               |||||||||||||||||||
Db         602 AAGAGTTTGTCAGTGGGAG 584
```

XX
XX
PI   Igarashi K, Kurokawa T, Nishizuka Y, Ono Y;
XX
DR   WPI; 1996-021913/03.
DR   P-PSDB; AAR94951.
XX
PT   New isolated rat protein kinase C - used to develop prods. for the
PT   study, diagnosis, prevention and treatment of diseases involving
PT   abnormal signal transduction
XX
PS   Example 2; Fig 2; 39pp; English.
XX
CC   A cDNA sequence (AAT18809) codes for a C-terminal portion (AAR94951)
CC   of type I (beta-1) rat brain protein kinase C (PKC), a protein that
CC   performs the transduction of extracellular signals into cells through
CC   the phosphorylation of proteins. It was obtd. from a rat brain-
CC   derived cDNA library by screening with probes (see also AAT18806-08)
CC   based on isolated peptides of PKC. A PstI fragment of the cDNA clone
CC   was used to screen a rat brain cDNA library, resulting in the
CC   identification of full-length sequences (see also AAQ74001 and AAT15006)
CC   coding for type I (AAR94953) and type II (AAR94764) PKC.
XX
SQ   Sequence 1177 BP; 318 A; 244 C; 277 G; 337 T; 1 other;
XX

Query Match   57.6%; Score 19; DB 17; Length 1177;
Best Local Similarity   100.0%; Pred. No. 53;
Matches   19; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy          15 AAGAGTTTGTCAGTGGGAG 33
               |||||||||||||||||||
Db         602 AAGAGTTTGTCAGTGGGAG 584
```

RESULT 14
AAN82021/c
ID   AAN82021 standard; DNA; 1802 BP.
XX
AC   AAN82021;
XX
DT   16-OCT-1990 (first entry)
XX
DE   Sequence encoding protein kinase C-III (PKC-III).
XX
KW   Protein kinase C; PKC; cancer; ds.
XX
OS   Rattus sp.
XX
PN   WO8801303-A.
XX
PD   25-FEB-1988.
XX
PF   13-AUG-1987; 87WO-0002005.
XX
PR   13-AUG-1986; 86US-0896476.
XX
PA   (GENE-) GENETICS INST INC.
XX
PI   Knopf JL;
XX
DR   WPI; 1988-064018/09.
DR   P-PSDB; AAP92021.
XX
PT   New DNA sequences coding for protein C enzyme -
PT   and new expressed polypeptide(s), useful for detecting tumour
PT   promoting activity of test cpds.
XX
PS   Claim 1; Fig 1; 36pp; English.
XX
CC   PKC product can be used in determining tumour promoting properties
CC   of a test sample, and probes derived from the sequence may be used
CC   in isolating polypeptides from other species.
XX
SQ   Sequence 1802 BP; 490 A; 438 C; 476 G; 397 T; 1 other;

```
Query Match          57.6%;     Score  19;   DB   9;   Length  1802;
Best Local Similarity  100.0%;     Pred. No. 58;
Matches    19;   Conservative   0;   Mismatches     0;   Indels     0;   Gaps     0;

Qy      15 AAGAGTTTGTCAGTGGGAG  33
           |||||||||||||||||||
Db    1604 AAGAGTTTGTCAGTGGGAG  1586
```

RESULT 15
AAN90750/c
ID   AAN90750 standard; DNA; 2600 BP.
XX
AC   AAN90750;
XX
DT   11-JAN-1990   (first entry)
XX
DE   Protein kinase C beta 1 cDNA.
XX
KW   Protein kinase C beta 1; cellular growth control; tumorigenesis.
XX
OS   Rat.
XX
PN   WO8907654-A.
XX
PD   24-AUG-1989.
XX
PF   09-FEB-1989;   89WO-US00462.
XX
PR   10-FEB-1988;   88US-0154206.
XX
PA   (PROG) PROGENICS PHARMACEUTICALS INC.
XX
PI   Housey GM;
XX
DR   WPI; 1989-263721/36.
DR   P-PSDB; AAP91373.
XX
PT   Screening for protein inhibitors and activators - by examining phenotypic
PT   changes of two cell lines having different protein prodn.
XX
PS   Disclosure; fig 1; 75pp; English.
XX
CC   Protein kinase C (PKC) beta 1 cDNA was sequenced from clone RP58
CC   isolated from a rat brain cDNA library corresp. to clone RP41. PKC is a
CC   Ca2+ - and phospholipid dependent ser/thr protein kinase important in
CC   cellular growth control.
XX
SQ   Sequence 2600 BP; 676 A; 642 C; 670 G; 612 T; 0 other;

```
Query Match          57.6%;     Score  19;   DB  10;   Length  2600;
Best Local Similarity  100.0%;     Pred. No. 62;
Matches    19;   Conservative   0;   Mismatches     0;   Indels     0;   Gaps     0;

Qy      15 AAGAGTTTGTCAGTGGGAG  33
           |||||||||||||||||||
Db    2033 AAGAGTTTGTCAGTGGGAG  2015
```

Search completed: December 16, 2002, 11:13:01
Job time : 940 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

Run on: December 16, 2002, 10:54:59 ; Search time 3176 Seconds
(without alignments)
302.391 Million cell updates/sec

Title: US-09-435-471B-9
Perfect score: 33
Sequence: 1 gcattctagtcgacaagagttgtcagtgggag 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455140287B residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

	1:	gb_ba:*
	2:	gb_htg:
	3:	gb_in:*
	4:	gb_om:*
	5:	gb_ov:*
	6:	gb_pat:*
	7:	gb_ph:*
	8:	gb_pl:*
	9:	gb_pr:*
	10:	gb_ro:*
	11:	gb_sts:*
	12:	gb_sy:*
	13:	gb_un:*
	14:	gb_vi:*
	15:	gb_vrt:*
	16:	em_ba:*
	17:	em_hum:*
	18:	em_fun:*
	19:	em_in:*
	20:	em_mu:*
	21:	em_om:*
	22:	em_or:*
	23:	em_ov:*
	24:	em_pat:*
	25:	em_ph:*
	26:	em_pl:*
	27:	em_ro:*
	28:	em_sts:*
	29:	em_un:*
	30:	em_vi:*
	31:	em_htg_hum:*
	32:	em_htg_inv:*
	33:	em_htg_mus:*
	34:	em_htg_other:*
	35:	em_htg_pln:*
	36:	em_htg_rod:*
	37:	em_htg_mam:*
	38:	em_vrt:*
	39:	em_sy:*
	40:	em_htgo_hum:*
	41:	em_htgo_mus:*
	41:	em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	% DB	ID	Description
1		23.2	70.3	180673	10	AC068627	AC068627 Mus Muscu
2	c	22	66.7	11260	1	AE014010	AE014010 Yersinia
3	c	22	66.7	203728	1	AJ414160	AJ414160 Yersinia
4	c	21.8	66.1	113451	2	AC023297	AC023297 Homo sapi
5	c	21.8	66.1	200724	9	AL139382	AL139382 Human DNA
6	c	21.6	65.5	160708	2	AC011962	AC011962 Homo sapi
7	c	21.6	65.5	163823	9	AC097372	AC097372 Homo sapi
8	c	21.4	64.8	199006	2	AC126835	AC126835 Rattus no
9	c	20.8	63.0	161308	2	AC122578	AC122578 Rattus no
10	c	20.8	63.0	179250	2	AC128864	AC128864 Rattus no
11	c	20.8	63.0	180035	2	AC117985	AC117985 Papio cyn
12	c	20.8	63.0	180783	2	AC116932	AC116932 Papio cyn
13	c	20.8	63.0	183698	10	AL606464	AL606464 Mouse DNA
14	c	20.8	63.0	209876	2	AL627315	AL627315 Mus muscu
15	c	20.6	62.4	169514	2	AC122449	AC122449 Mus muscu
16	c	20.4	61.8	72781	2	AC018417	AC018417 Homo sapi
17	c	20.4	61.8	98240	9	AC006021	AC006021 Homo sapi
18	c	20.4	61.8	109244	2	AC109077	AC109077 Rattus no
19	c	20.4	61.8	109395	2	AC013742	AC013742 Homo sapi
20	c	20.4	61.8	158987	2	AC108713	AC108713 Homo sapi
21	c	20.4	61.8	180189	9	AL365510	AL365510 Human DNA
22	c	20.4	61.8	184090	2	AC009564	AC009564 Homo sapi
23	c	20.4	61.8	198481	9	AC022389	AC022389 Homo sapi
24	c	20.2	61.2	40286	3	CEC05E7	CEC05E7 Caenorhabdi
25	c	20.2	61.2	124278	2	AC104618	AC104618 Homo sapi
26	c	20.2	61.2	135706	2	AC102862	AC102862 Mus muscu
27	c	20.2	61.2	161421	2	AC068114	AC068114 Homo sapi
28	c	20.2	61.2	169226	9	AL590669	AL590669 Human DNA
29	c	20.2	61.2	179371	2	AC103130	AC103130 Rattus no
30		20	60.6	892	3	AF302260	AF302260 Biomphala
31		20	60.6	1496	3	UEU408960	UEU408960 Uncultur
32		20	60.6	24707	3	Z66522	Z66522 Caenorhabdi
33		20	60.6	178893	2	AC126681	AC126681 Rattus no
34		20	60.6	181911	2	AC098105	AC098105 Rattus no
35		20	60.6	193006	2	AC120599	AC120599 Rattus no
36		19.8	60.0	401	6	AX270035	AX270035 Sequence
37		19.8	60.0	401	6	AX270036	AX270036 Sequence
38		19.8	60.0	401	6	AX271566	AX271566 Sequence
39		19.8	60.0	401	6	AX271567	AX271567 Sequence
40		19.8	60.0	2251	3	AY119516	AY119516 Drosophil
41		19.8	60.0	32814	8	Z99295	Z99295 S.pombe chr
42		19.8	60.0	99895	2	AC114858	AC114858 Rattus no
43		19.8	60.0	106551	2	AC017423	AC017423 Drosophil
44		19.8	60.0	110000	9	AF491780_08	Continuation (9 of
45		19.8	60.0	137009	9	AC009970	AC009970 Homo sapi

ALIGNMENTS

RESULT 1
AC068627/c
LOCUS   AC068627   180673 bp   DNA   linear   ROD 15-MAY-2002
DEFINITION   Mus musculus Chromosome 5 RP23-389F6, complete sequence.
ACCESSION   AC068627
VERSION   AC068627.16 GI:13027370
KEYWORDS   HTG.
SOURCE   Mus musculus
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostom1;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 180673)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
               Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
               Gordon,M., Goltz,J.S. and Kucherlapati,R.

```
TITLE      High Throughput Mouse Sequencing
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 180673)
AUTHORS    Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
           Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
           Gordon,M., Goltz,J.S. and Kucherlapati,R.
JOURNAL    Direct Submission
           Submitted (06-MAY-2000) Department of Molecular Genetics, Albert
           Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
           Bronx, NY 10461, USA
REFERENCE  3  (bases 1 to 180673)
AUTHORS    Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
           Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
           Gordon,M., Goltz,J.S. and Kucherlapati,R.
JOURNAL    Direct Submission
           Submitted (21-FEB-2001) Department of Molecular Genetics, Albert
           Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
           Bronx, NY 10461, USA
TITLE      Direct Submission
           ------- Genome Center:
           4  (bases 1 to 180673)
REFERENCE  On Feb 21, 2001 this sequence version replaced gi:11276096.
AUTHORS    Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
           Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
           Gordon,M., Goltz,J.S. and Kucherlapati,R.
JOURNAL    Direct Submission
           Submitted (24-FEB-2001) Department of Molecular Genetics, Albert
           Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
           Bronx, NY 10461, USA
TITLE      High Throughput Mouse Sequencing
JOURNAL    Unpublished

COMMENT    On Feb 21, 2001 this sequence version replaced gi:11276096.
           Contact:  hpgc@mendel.mgh.harvard.edu
           Web site: http://www.hpcg.org/Sequence/mouse.html
           Center Code:  HPGC
           Harvard Partners Genome Center:
           ------- Genome Center:

           CLONE LENGTH: This sequence represents the entire insert of this
           clone unless otherwise noted. If there are overlapping clones, the
           overlaps are noted in the beginning and end of the Features
           listing.

           ANNOTATION OF FEATURES:
           STSs are identified using ePCR (Genome Res. 7:541-550).
           Repeats are identified using RepeatMasker (A. Smit and P. Green,
           unpublished.) for Human and Mouse sequences.
           Genes and Regions of sequence similarity are identified by BLAST
           (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST
           and cDNA sequences in Unigene. Genes demonstrate at least two exons
           flanked by consensus splice sites that maintain sequence continuity
           across the splice junctions. Sequences that are not identical
           matches are annotated as similar.

           SEQUENCING READ COVERAGE: Attempts are made to complete double
           stranded sequence for all regions. All sequence is completed to a
           standard of coverage with a minimum of 3 reads with no ambiguities.
           If the sequence coverage for a region does not meet this standard,
           it is indicated in the annotation as Low Coverage. Low coverage
           linkages are verified by PCR product size verification or
           verification of forward and reverse reads from clones which span
           the low coverage area.

           QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
           standards - estimated average error rate is less than 1 per 10,000
           bases using the Consed average quality parameters. Regions that do not
           meet this requirement are annotated as Low Quality.

           ------------------Summary Statistics
           Center project name: ACQ
           Sequencing vector: pUC18;  L08752
           Chemistry: Dye-terminator Big Dye: 100%
           Assembly program: Phrap version 0.90319
           Contig length: 180673
           Fraction of Phrap value < 40: 0.00121
           Error Rate in Consed: .01 per 10,000 bases
           Number of N's in consensus: 0
```

```
         -------------- Distribution of Quality < 40 Bases:

           1000|
            900|
            800|
            700|
          # 600|
      bases 500|
            500|
            400|
            400|
            300|
            200|
            100|
              0|
                 5    10   15   20   25   30   35   40
                        Phrap Value Range

FEATURES             Location/Qualifiers
     source          1..180673
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-389F6"
     misc_feature    1..45796
                     /note="Overlap with Clone RP23-327P24, AC084217"
     repeat_region   160..261
                     complement(274..410)
     repeat_region   2322..2361
                     /rpt_family="MTB"
     repeat_region   /rpt_family="CT-rich"
                     complement(658..694)
                     /rpt_family="B1_MM"
     repeat_region   3396..3424
                     /rpt_family="(TA)n"
     repeat_region   3425..3748
                     /rpt_family="(CAAA)n"
     repeat_region   complement(695..840)
                     /rpt_family="B3"
     repeat_region   3749..3771
                     /rpt_family="ORR1A2"
     repeat_region   1361..1558
                     /rpt_family="B3"
     repeat_region   4113..4161
                     /rpt_family="(CAAA)n"
     repeat_region   1635..1667
                     /rpt_family="B3"
     repeat_region   complement(1922..2321)
                     /rpt_family="B3"
     repeat_region   1667..1921
                     /rpt_family="(TC)n"
     repeat_region   5096..5125
                     /rpt_family="AT_rich"
     repeat_region   5932..6099
                     /rpt_family="(TAA)n"
     repeat_region   complement(4799..5020)
                     /rpt_family="B3A"
     repeat_region   6102..6137
                     /rpt_family="(CA)n"
     repeat_region   6962..7044
                     /rpt_family="B3"
     repeat_region   7119..7151
                     /rpt_family="B3"
     repeat_region   complement(7223..7369)
                     /rpt_family="(CCCCCT)n"
     repeat_region   complement(7518..7668)
                     /rpt_family="B1_MM"
     repeat_region   7677..8041
                     /rpt_family="URR1A"
     repeat_region   complement(8046..8104)
                     /rpt_family="RMER10B"
     repeat_region   8429..78556
                     /rpt_family="URR1B"
     STS             /standard_name="D5Mit286"
```

```
RESULT 2
LOCUS       AE014010/c       11260 bp    DNA    linear    BCT 26-JUL-2002
DEFINITION  Yersinia pestis KIM section 410 of 415 of the complete genome.
ACCESSION   AE014010
VERSION     AE014010.1  GI:21961093
KEYWORDS
SOURCE      Yersinia pestis KIM
  ORGANISM  Yersinia pestis KIM
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 11260)
  AUTHORS   Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
            Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
            Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
            Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
            Blattner,F.R. and Perry,R.D.
  TITLE     Genome Sequence of Yersinia pestis KIM
  JOURNAL   J. Bacteriol. 184 (16), 4601-4611 (2002)
  PUBMED    12142430
REFERENCE   2  (bases 1 to 11260)
  AUTHORS   Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
            Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
            Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
            Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
            Blattner,F.R. and Perry,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
            Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1..11260
                     /organism="Yersinia pestis KIM"
                     /strain="KIM"
                     /db_xref="taxon:187410"
     gene            126..458
                     /gene="y4082"
     CDS             126..458
                     /gene="y4082"
                     /function="unknown"
                     /note="residues 24 to 110 of 110 are 32.95 pct identical
                     to residues 9 to 96 of 96 from GenPept : >emb|CAD08729.1|
                     (AL627266) hypothetical protein [Salmonella enterica
                     subsp. enterica serovar Typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical"
                     /protein_id="AAM87625.1"
                     /db_xref="GI:21961094"
                     /translation="MAVKKITPIAQRTVDSMKKIIFFIMPFFILLGYSHAAEKYPADI
                     TAYLNVADDCQYFSGGEWDSTLPKEROIEIEKKVNVTCSKARSLQEKLSVKYKKRQDLL
                     DVINDVDF"
     gene            990..1466
                     /gene="y4083"
     CDS             990..1466
                     /gene="y4083"
                     /function="unknown"
                     /note="residues 71 to 155 of 158 are 30.33 pct identical
                     to residues 102 to 190 of 195 from GenPept :
                     >emb|CAB51120.1| (AL109962) putative secreted protein
                     [Streptomyces coelicolor A3(2)]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical"
                     /protein_id="AAM87626.1"
                     /db_xref="GI:21961095"
                     /translation="MTNRMKALMLVTHLAVFGAGTGTLGIYILPLLTAETNATAPEIQA
                     VKQSAXYTGFESRNQKGSDAVHWAEGKLVYSEHIEFAFEGEIAPGPDYKIYLTKTQADD
                     KESFLKIKDDAKLLGDLKNFGNFKKTLPAGVNPDDYTTVQIWCEFFSQFIGSASYK"
     gene            complement(1635..1991)
                     /gene="y4084"
     CDS             complement(1635..1991)
                     /gene="y4084"
                     /function="unknown"
                     /note="residues 3 to 118 of 118 are 61.20 pct identical to
                     residues 5 to 120 of 120 from E. coli K12 : B3602;
                     residues 1 to 118 are 100.00 pct identical to
                     residues 1 to 118 of 118 from GenPept : >emb|CAC93519.1|
                     (AJ414160) conserved hypothetical protein [Yersinia
                     pestis]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM87627.1"
                     /db_xref="GI:21961096"
                     /translation="MKEHAEIKRLSDMLDALNHKDPTVIQQGNVELITQHMKEKEKLA
                     AEIQRLKEVRVKNLSAEAQKLAQLPFSRAITKKEQADMGTLKRAVRGIVVHPMTALG
                     REMGLKEVTGYAKRAF"
     gene            complement(2240..2794)
                     /gene="mtlR"
     CDS             complement(2240..2794)
                     /gene="mtlR"
                     /function="regulator; degradation of small molecules;
                     Carbon compounds"
                     /note="residues 2 to 174 of 184 are 77.45 pct identical to
                     residues 22 to 191 of 195 from E. coli K12 : B3601;
```

```
     repeat_region   8466..8555
                     /rpt_family="(TA)n"
     repeat_region   8785..8853
                     /rpt_family="(TA)n"
     repeat_region   9945..9978
                     /rpt_family="(TA)n"
     repeat_region   10102..10298
                     /rpt_family="(GGGGA)n"
     repeat_region   10610..10633
                     /rpt_family="MLT1A2"
     repeat_region   10457..10529
                     /rpt_family="RMER20"
     repeat_region   11378..11518
                     /rpt_family="(CA)n"
     repeat_region   12007..12256
                     /rpt_family="(CA)n"
     repeat_region   12620..12669
                     /rpt_family="(CA)n"
     repeat_region   complement(12672..12796)
                     /rpt_family="T-rich"
     repeat_region   12794..12936
                     /rpt_family="B1_MM"
     repeat_region   12959..13132
                     /rpt_family="B3"
     repeat_region   13153..13509
                     /rpt_family="B2_Mm2"
     repeat_region   14402..14629
                     /rpt_family="L1MEd"
     repeat_region   14781..14830
                     /rpt_family="B3"
     repeat_region   complement(16154..16310)
                     /rpt_family="B2_Mm2"
     repeat_region   complement(16027..16083)
                     /rpt_family="B2_Mm2"
     repeat_region   15969..16010
                     /rpt_family="PB1"
     repeat_region   complement(15773..15888)
                     /rpt_family="PB1"
     repeat_region   complement(14964..15073)
                     /rpt_family="B1_MM"
     repeat_region   complement(14830..?)
                     /rpt_family="(TG)n"
     repeat_region   complement(16325..16472)
                     /rpt_family="B2_Mm1"
```

```
QY          1 GCATTCTAGTCGACAAGAGTTGTCAGT 28
              |||||||||||| |||||||||||| |
Db     106672 GCATTCTAGTCCACAAGAGTTGTGTTATT 106645

Query Match   70.3%;  Score 23.2;  DB 10;  Length 180673;
Best Local Similarity  89.3%;  Pred. No. 6.7;
Matches  25;  Conservative  0;  Mismatches  3;  Indels  0;  Gaps  0;

QY          1 GCATTCTAGTCGACAAGAGTTGTCAGT 28
              |||||||||||| |||||||||||| |
Db     106672 GCATTCTAGTCCACAAGAGTTGTTTATT 106645
```

residues 1 to 181 of 184 are 75.69 pct identical to
residues 21 to 198 of 198 from GenPept :
>gbl|AAD45387.1|AF166095_3 (AF166095) mtl operon repressor
[Klebsiella pneumoniae]"
/codon_start=1
/trans_table=11
/product="repressor for mtl"
/protein_id="AAM87628.1"
/db_xref="GI:21961097"
/translation="MIEKKQAFENRVLEHLNAGKTVRSFLMAAVDLLEALNILVVQI
FRKDYAVKYAVEPLLVGDPGLSELSVRLKLVYALGVITRHEYEDAEGLLMALREELNH
DGYEYRFTDDEILGPPGELHCVDELPPVPTFLRPDEADASLIAMQRQRYQOMVRSTMV
LSTPELLSRISVKQVSKLSPLSQG"

gene
/gene="mtlD"
/note="y4086"
complement(3001..4269)

CDS
complement(3001..4269)
/gene="mtlD"
/function="enzyme; degradation of small molecules; Carbon
compounds"
/note="residues 36 to 417 of 422 are 75.65 pct identical
to residues 1 to 382 of 382 from E. coli K12 : B3600"
/codon_start=1
/trans_table=11
/product="mannitol-1-phosphate dehydrogenase"
/protein_id="AAM87629.1"
/db_xref="GI:21961098"
/translation="MAALPTCNANYLAYKQVITARAVGRKWPFYSKVTNMKALHFGAG
NIGRGFIGKLLADAGAQLTFADVNQPLLDELNKRKRYQVNVVGEQRAVEVEKVNYSAVN
SGSPEVVALIAEADIVTTAVGPQILARIAATVAQGLITRHQQGNTRPLNIIACENMVR
GTSQLKOHVFAALSEDEQIWEOHVGFVDSAVDRIVPPSEAGSTDILAVTIVETFSEWI
VDGTOFKGOPPEIVGMELTDNLMAFVERKLFTLNGHATTAYLGQLAGHQTIRDAILD
PAVROTVKGAMEESGAVLIKRYAFPDQKHAAYINKILSRFENPYLHDDVERVGRDPLR
KLSAGDRLIKPLLGTLEYQLPHDSLVTGIAAAMSYRSEQDPQAQELVTLLAQLGPKAA
LAQISGLPADSEVVEQGAVSYYNAMQQKLAH"

gene
/gene="mtlA"
/note="y4087"
complement(4301..6232)

CDS
complement(4301..6232)
/gene="mtlA"
/function="transport; transport of small molecules;
carbohydrates, organic acids, alcohols"
/note="residues 3 to 640 of 643 are 81.66 pct identical to
residues 2 to 636 of 637 from E. coli K12 : B3599;
residues 3 to 640 of 638 from GenPept : >gbl|AAL22544.1|
(AE008871) PTS family, mannitol-specific enzyme IIABC
components [Salmonella typhimurium LT2]"
/codon_start=1
/trans_table=11
/product="PTS system mannitol-specific enzyme IIABC
components"
/protein_id="AAM87630.1"
/db_xref="GI:21961099"
/translation="MFSPDAKVRVQNFGGRFLSNMVMPNIGAFIAWGIITALFIPTGWL
PNETLAKVLVGPMITVLLPLLIGFTGRLVGGDRGVGVGAITIMGVIVGADMPMFLGAM
IVGPLGGWAIKIHFDRRWVEKRIKSGFEMLVHNVRFSSGIIGNLALLALLAFMAIGFLEVLSK
GLAYGVDWVQNNLLPLASIFVEPAKILFLNNAIHNGISPSLCGVQOAAEFGSKSIFFLI
EANGPGLGVLMAYMFFGNGNAKOSAGAAIHFEGGIHEIYFPYVLMNPRLLAVIL
GGMTGVFTLTLNGGLVSASPGSILATLAMTPKGAYFANINAVATAFAVSFVVSAIL
LKSSAKIDDEGLEGATRRMODMKAQSKGVQAANAAAGDLSTVRKIIVACDAGMGS
SAMGAGVLRKVYQDAGLIKHIAVINCAINNLPEDVDLVITHRDLTERAMRHAPAQHIS
LTNFLDSQLYNSLTAELLQASQLVDNRKVQKVIEKLDDSFEAFEDNLFKLGAENIFLN
QHATAKEQAIRFAGEQLVKGGYVEPEVVEAMLDREKLSSTYLGESIAVPHGTIEAKDR
VLKTGVVFCQVPEGYRFGEEDDVARLVIGIAARNEHIQVITSLINALDDDSVIARL
SKTTSVQEVLDLLGGKTSA"

gene
/gene="y4088"
complement(6882..7607)

CDS
complement(6882..7607)
/gene="y4088"
/note="residues 14 to 226 of 241 are 53.05 pct identical
to residues 49 to 261 of 276 from E. coli K12 : B3554;
residues 14 to 226 of 241 are 52.11 pct identical to

residues 9 to 221 of 236 from GenPept : >gbl|AAL22507.1|
(AE008869) putative outer membrane lipoprotein [Salmonella
typhimurium LT2]"
/codon_start=1
/trans_table=11
/product="hypothetical protein"
/protein_id="AAM87631.1"
/db_xref="GI:21961100"
/translation="MHHIFALNTVRNPRWLAPLLAFIVMFLVYLVTACGDKELEQRKAFID
YLQNTVMRSGMKLPTLSEDQKGFGPGYSVDYAIILVTFEQSQOLSKSVDASLVPAIEQINQ
IRVAQDYLSRDALQQSAGAINLLVQQIRTAKRAQADSAQQAALKLPDDLKVFNRAFDN
IVTOPANVIIPAYPVVSSFVODLQVGVGDFPLQOGGTQVTFNNGGIQFQTPQQAAQYNTM
MSNLVAKIYPEMMAAQKRVMKVMQ"
complement(8074..10143)
/gene="glyS"
/note="y4089"

CDS
complement(8074..10143)
/gene="glyS"
/function="enzyme; aminoacyl tRNA synthetases, tRNA
modification"
/note="residues 1 to 689 of 689 are 85.63 pct identical to
residues 1 to 689 of 689 from E. coli K12 : B3559;
residues 1 to 689 of 689 are 86.50 pct identical to
residues 1 to 689 of 689 from GenPept : >gbl|AAL22514.1|
(AE008869) glycine tRNA synthetase, beta subunit
[Salmonella typhimurium LT2]"
/codon_start=1
/trans_table=11
/product="glycine tRNA synthetase, beta subunit"
/protein_id="AAM87632.1"
/db_xref="GI:21961101"
/translation="MTQQTFLVEIGTEELPPKRALRSLAESFAANFTAELDNANLSHGE
VSWYAPRRLAVKVYANLSAAOADREVEKRGPAIAQAPDAEGKPSKAAEGWARGCGTTV
DQAERLVTDKGEWLLYRAHVKGOPAQLLLAGMVNTALSKLPIPKLMRWGDKETQFVRP"

```
Qy    1 GCATTCTAGTCGCACAACAGTTTGTCAGTGG 30
        |||||-|-|||||-|||||| ||||||| |||
Db 8678 GCATTATCGTCGAAAAGAATTTGTCACTGG 8649
```

Query Match   66.7%;   Score 22;   DB 1;   Length 11260;
Best Local Similarity   83.3%;   Pred. No. 29;
Matches   25;   Conservative   0;   Mismatches   5;   Indels   0;   Gaps   0;

RESULT 3
AJ414160/c
LOCUS   AJ414160   203728 bp   DNA   linear   BCT 06-JUN-2002
DEFINITION Yersinia pestis strain CO92 complete genome; segment 20/20.
ACCESSION   AJ414160 AL590842 AL590844
VERSION   AJ414160.1 GI:15981860
KEYWORDS   .
SOURCE   Yersinia pestis
ORGANISM   Yersinia pestis
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1 (bases 1 to 203728)
AUTHORS   Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,
            Prentice,M.B., Sebaihia,M., James,K.D., Churcher,C., Mungall,K.L.,
            Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,
            Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,
            Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,
            Karlyshev,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,
            Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G.
TITLE   Genome sequence of Yersinia pestis, the causative agent of plague
JOURNAL   Nature 413 (6855), 523-527 (2001)
MEDLINE   21470413
PUBMED   11586360
REFERENCE   2 (bases 1 to 203728)
AUTHORS   Parkhill,J.
TITLE   Direct Submission
JOURNAL   Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E'mail: parkhill@sanger.ac.uk

```
COMMENT     Notes:
            Details of Y. pestis sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
FEATURES             Location/Qualifiers
     source          1..203728
                     /organism="Yersinia pestis"
                     /strain="CO92"
                     /db_xref="taxon:632"
                     /note="biovar: Orientalis"
     gene            complement(187..420)
                     /gene="YPO3952"
     CDS             complement(187..420)
                     /gene="YPO3952"
                     /note="no significant database hits to the full length
                     CDS. Similar to the N-terminal regions of YPO0639 (59.5%
                     identity in 79 aa overlap) and YPO1258 (59.5% identity in
                     79 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC93414.1"
                     /db_xref="GI:15981861"
                     /db_xref="SPTREMBL:Q8ZA66"
                     /translation="MSLSLLGIYLYINLSRDNLRPEEYERIVSAYAAWTRVCREYEFND
                     GYHANHYIINMRERQVHHTQASSLGPREKDLE"
     gene            complement(602..1105)
                     /gene="YPO3953"
     CDS             complement(602..1105)
                     /gene="YPO3953"
                     /EC_number="2.7.1.12"
                     /note="Similar to Escherichia coli thermosensitive
                     gluconokinase IdnK SW:IDNK_ECOLI (P39208) (187 aa) fasta
                     scores: E(): 0, 59.1% id in 159 aa, and to Escherichia
                     coli thermoresistant gluconokinase GntK SW:GNTK_ECOLI
                     (P46859) (175 aa) fasta scores: E(): 0, 59.2% id in 152
                     aa. Similar to YPO2540 (66.3% identity in 160 aa overlap)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative gluconokinase"
                     /protein_id="CAC93415.1"
                     /db_xref="GI:15981862"
                     /db_xref="SPTREMBL:Q8ZA65"
                     /translation="MAQQSIIVMGVSGSGKTTVGEAVAAQIHAKFIDGDDLHPRANIQ
                     KMGSGQPLNDADRMPWLERLSDAAYSLSHKNETGIIVCSALKRCYRDRLREGNQGMVF
                     LYLKGNFDVIMARLQARSGHPMSDLLRSQFDALEEPGDEPDVICVDIDTDIDEVVR
                     RCVGALA"
     misc_feature    complement(1055..1078)
                     /gene="YPO3953"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)."
     gene            1322..2638
                     /gene="YPO3954"
     CDS             1322..2638
                     /gene="YPO3954"
                     /note="gntM; gntT; usgA"
                     /note="Similar to Escherichia coli high-affinity gluconate
                     transporter GntT SW:GNTT_ECOLI (P39835) (437 aa) fasta
                     scores: E(): 0, 84.5% id in 438 aa, and to Escherichia
                     coli Gnt-II system L-idonate transporter IdnT
                     SW:IDNT_ECOLI (P39344) (439 aa) fasta scores: E(): 0,
                     61.2% id in 438 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="putative gluconate permease"
                     /protein_id="CAC93416.1"
                     /db_xref="GI:15981863"
                     /db_xref="SPTREMBL:Q8ZA64"
                     /translation="MPLIIVAVGVAMLLLMIRYKLNGFISLILVALAVGIMQGMPID
                     KVVGSIKAGVGGTLGLSLALIMGFAAMLGKLLADCGGAQRIATTLIEKFGDKHIQWAVV
                     LTGFTVGRALFYEVGFLGVLLPLVFSIAASARIPLLVGVGPMAAALSVTHGFLPHPHPG
                     TAIAIFNADMGKFLLGTLLIGTLLAIPIVILAGPVYARFLKGIDKPVPEGIYNKIFTDKE
                     MPFGVVSTALVPVILMALRAVAEMILPKHSLLPYAEFFGPDVIAFTLIAVLLAIFT
                     FGLNRGRSVESVWETLSDSIKIIAMMLLVIGGGGAFKQVLVDSGIEQYIAGMGMNSGA
```

```
     misc_feature    1322..2632
                     /gene="YPO3954"
                     /note="Pfam match to entry PF02447 GntP_permease, , score
                     926.00, E-value 1.1e-274"
     misc_feature    1364..1432
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    1469..1537
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    1640..1708
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    1727..1786
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    1844..1912
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    1979..2047
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    2090..2158
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    2216..2284
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    2356..2434
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    2453..2521
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     misc_feature    2564..2632
                     /gene="YPO3954"
                     /note="one of 11 probable transmembrane helices predicted
                     for YPO3954 by TMHMM2.0"
     gene            complement(2730..3725)
                     /gene="YPO3955"
     CDS             complement(2730..3725)
                     /gene="YPO3955"
                     /note="Similar to Escherichia coli gluconate utilization
                     system Gnt-I transcriptional repressor GntR SW:GNTR_ECOLI
                     (P46860) (331 aa) fasta scores: E(): 0, 79.1% id in 325
                     aa, and to Vibrio cholerae gluconate utilization system
                     Gnt-I transcriptional repressor VC0289 TR:Q9KV68
                     (EMBL:AE004117) (332 aa) fasta scores: E(): 0, 52.5% id in
                     324 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="gluconate utilization system Gnt-I
                     transcriptional repressor"
                     /protein_id="CAC93417.1"
                     /db_xref="GI:15981864"
                     /db_xref="SPTREMBL:Q8ZA63"
                     /translation="MKKRPVLQDVADMVGVTKMTVSRYLRNPEQVSAVLQERIALAL
                     DELGYIPNRAPDILSNATSHAICGVLLPSLTNQVPAEVLRGIESVTDAHGYQTMLAHYG
                     YOKEREORLTSLLSYNIDGLILSERHHTARTLKMIEVAGIPVIELMDGISPCLDLAV
                     GFNNFEAAROWTQQIIAKGHRVYVYFGARODERIVIKQQGYEQOAMRESGLVPMSVMTS
```

```
                     RSSYSAGAELLREAQVSYPQIDISIFCTNDDLAIGAVFFCCQRQGLSIPHDMAIAGFHG
                     HDIGOSMVPKLASVLIPRERMGQIGAERLLARLRGEGVTPQMVDVGFTILPGGSI"
misc_feature         complement(2823..3542)
                     /gene="YPO3955"
                     /note="Pfam match to entry PF00532 Peripla_BP_like,
                     Periplasmic binding proteins and sugar binding domain of
                     the LacI family." score 145.80, E-value 7.6e-40"
misc_feature         complement(3633..3704)
                     /gene="YPO3956"
                     /note="Pfam match to entry PF00356 lacI, Bacterial
                     regulatory proteins, lacI family, score 30.50, E-value
                     4.4e-07"
misc_feature         complement(3648..3704)
                     /gene="YPO3955"
                     /note="PS00356 Bacterial regulatory proteins, lacI family
                     signature."
gene                 complement(3854..4549)
                     /gene="YPO3956"
CDS                  complement(3854..4549)
                     /gene="YPO3956"
                     /note="Similar to Escherichia coli hypothetical protein
                     YhhW SW:YHHW_ECOLI (P46852) (231 aa) fasta scores: E(): 0,
                     77.5% id in 231 aa, and to Synechocystis sp hypothetical
                     protein SLL1773 TR:P73623 (EMBL:D90008) (232 aa) fasta
                     scores: E(): 0, 50.4% id in 234 aa"
                     /codon_start=1
                     /transl_table=11

Qy       1 GCATTCTTCACAAGAGTTTGTCACGTGG  30
           |||||  ||| ||||| |||||| |||| |||
Db 141825 GCATTATCGTCGAAAAGAATTTGTCACTGG 141796


Query Match      66.7%;   Score 22;   DB 1;   Length 203728;
Best Local Similarity  83.3%;   Pred. No. 23;
Matches   25;   Conservative   0;   Mismatches   5;   Indels   0;   Gaps   0;


RESULT 4
AC023297/c
LOCUS       AC023297   113451 bp    DNA   linear   HTG  03-MAR-2000
DEFINITION  Homo sapiens clone RP11-21H9, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC023297
VERSION     AC023297.3  GI:7144965
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113451)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 113451)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,P., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,J.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Minova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,P., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigllio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:7139552.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -----------------------------------------------------
            -- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -----------------------------------------------------
            -- Project Information
            Center project name: L4020
            Center clone name: 21_H_9
            -----------------------------------------------------
            -- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 106693 bases at least Q40
            Consensus quality: 109791 bases at least Q30
            Consensus quality: 111178 bases at least Q20
            Insert size: 138000; agarose-fp
            Insert size: 112251; sum-of-contigs
            Quality coverage: 3.5 in Q20 bases; agarose-fp
            Quality coverage: 4.3 in Q20 bases; sum-of-contigs
            *
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *         1     2264:  contig of 2264 bp in length
            *      2265     2364:  gap of 100 bp
            *      2365     6779:  contig of 4415 bp in length
            *      6780     6879:  gap of 100 bp
            *      6880    11332:  contig of 4453 bp in length
            *     11333    11432:  gap of 100 bp
            *     11433    17335:  contig of 5903 bp in length
            *     17336    17435:  gap of 100 bp
            *     17436    23308:  contig of 5873 bp in length
            *     23309    23408:  gap of 100 bp
            *     23409    28598:  contig of 5190 bp in length
            *     28599    28698:  gap of 100 bp
            *     28699    36183:  contig of 7485 bp in length
            *     36184    36283:  gap of 100 bp
            *     36284    44528:  contig of 8246 bp in length
            *     44529    44628:  gap of 100 bp
            *     44630    53672:  contig of 9043 bp in length
            *     53673    53772:  gap of 100 bp
            *     53773    62948:  contig of 9176 bp in length
            *     62949    63048:  gap of 100 bp
            *     63049    75673:  contig of 12625 bp in length
            *     75674    75773:  gap of 100 bp
            *     75774    92375:  contig of 16602 bp in length
            *     92376    92475:  gap of 100 bp
            *     92476   113451:  contig of 20976 bp in length
FEATURES             Location/Qualifiers
source          1..113451
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-21H9"
                /clone_lib="RPCI-11 Human Male BAC"
misc_feature    1..2264
                /note="assembly_fragment"
misc_feature    2365..6779
                /note="assembly_fragment"
misc_feature    6880..11332
                /note="assembly_fragment"
```

```
misc_feature    11433..17335
                /note="assembly_fragment"
misc_feature    17436..23308
                /note="assembly_fragment"
misc_feature    23409..28598
                /note="assembly_fragment"
misc_feature    28699..36183
                /note="assembly_fragment"
                /note="assembly_fragment
                clone_end:T7
                vector_side:left"
misc_feature    36284..44529
                /note="assembly_fragment"
misc_feature    44630..53672
                /note="assembly_fragment"
misc_feature    53773..62948
                /note="assembly_fragment
                clone_end:SP6
                vector_side:right"
misc_feature    63049..75673
                /note="assembly_fragment"
misc_feature    75774..92375
                /note="assembly_fragment"
misc_feature    92476..113451
                /note="assembly_fragment"
BASE COUNT     34163 a   23042 c   23371 g   31675 t   1200 others
ORIGIN

Qy        1 GCATTCTTAGTGTGCACAAGAGTTTGTCACGTGGGAG 33
            ||||| |||||||||||||||||||| | |||||
Db    48392 GAATTCATGTTGACAAGAGTTTGGCTTTGGGAG 48360

Query Match      66.1%;  Score 21.8;  DB 2;  Length 113451;
Best Local Similarity  78.8%;  Pred. No. 30;
Matches   26;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

RESULT 5
LOCUS       AL139382              200724 bp    DNA    linear    PRI 30-NOV-2000
AL139382/c
DEFINITION  Human DNA sequence from clone RP11-86N24 on chromosome 13, complete
            sequence.
ACCESSION   AL139382
VERSION     AL139382.12  GI:11492233
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200724)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: Clonerequest@sanger.ac.uk
COMMENT     On Nov 30, 2000 this sequence version replaced gi:11322785.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
```

```
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-86N24 is from the library RPCI-11.1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-86N24 The true
            right end of clone RP11-117l13 is at 16379 in this sequence.
FEATURES             Location/Qualifiers
     source          1..200724
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-86N24"
                     /clone_lib="RPCI-11.1"
     misc_feature    49203
                     /note="Forced join. Assembly confirmed by restriction
                     digest."
BASE COUNT     60078 a   41888 c   40178 g   58580 t
ORIGIN

Qy        1 GCATTCTTAGTGTGCACAAGAGTTTGTCACGTGGGAG 33
            ||||| |||||||||||||||||||| | |||||
Db   108477 GAATTCATGTTGACAAGAGTTTGGCTTTGGGAG 108445

Query Match      66.1%;  Score 21.8;  DB 9;  Length 200724;
Best Local Similarity  78.8%;  Pred. No. 29;
Matches   26;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

RESULT 6
LOCUS       AC011962              160708 bp    DNA    linear    HTG 06-MAY-2001
AC011962/c
DEFINITION  Homo sapiens chromosome 4 clone RP11-166K23 map 4, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC011962
VERSION     AC011962.3  GI:8516050
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160708)
  AUTHORS   Birten,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-166K23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160708)
  AUTHORS   Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkly,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C. Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGuk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 14, 2000 this sequence version replaced gi:7630664.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------------------   Genome Center
            Center: Whitehead Institute/MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
```

```
      Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
      Center project name: L2589
      Center clone name: 166_K_23
----- Summary Statistics
      Sequencing vector: M13; M77815; 100% of reads
      Chemistry: Dye-terminator Big Dye; 100% of reads
      Assembly program: Phrap; version 0.960731
      Consensus quality: 148088 bases at least Q40
      Consensus quality: 154421 bases at least Q30
      Consensus quality: 157062 bases at least Q20
      Insert size: 151000; agarose-fp
      Insert size: 159408; sum-of-contigs
      Quality coverage: 4.6 in Q20 bases; agarose-fp
      Quality coverage: 4.4 in Q20 bases; sum-of-contigs
------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*        1        2140: contig of 2140 bp in length
*     2141        2240: gap of      100 bp
*     2241        5077: contig of 2837 bp in length
*     5078        5177: gap of      100 bp
*     5177       10215: contig of 5038 bp in length
*    10216       10315: gap of      100 bp
*    10316       14680: contig of 4365 bp in length
*    14681       14780: gap of      100 bp
*    14781       21149: contig of 6369 bp in length
*    21150       21249: gap of      100 bp
*    21250       26493: contig of 5244 bp in length
*    26494       26593: gap of      100 bp
*    26594       36585: contig of 9992 bp in length
*    36586       36685: gap of      100 bp
*    36686       43665: contig of 6980 bp in length
*    43666       43765: gap of      100 bp
*    43766       55870: contig of 12105 bp in length
*    55871       55970: gap of      100 bp
*    55970       66656: contig of 10686 bp in length
*    66657       66756: gap of      100 bp
*    66756       87066: contig of 20310 bp in length
*    87067       87166: gap of      100 bp
*    87167      109119: contig of 21953 bp in length
*   109120      109219: gap of      100 bp
*   109220      135181: contig of 25962 bp in length
*   135182      135281: gap of      100 bp
*   135282      160708: contig of 25427 bp in length.
FEATURES             Location/Qualifiers
     source          1..160708
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-166K23"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2140
                     /note="assembly_fragment"
     misc_feature    2241..5077
                     /note="assembly_fragment"
     misc_feature    5178..10215
                     /note="assembly_fragment"
     misc_feature    10316..14680
                     /note="assembly_fragment"
                     /clone_end:SP6
                     /vector_side:right"
     misc_feature    14781..21149
                     /note="assembly_fragment"
     misc_feature    21250..26493
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:right"
                     /note="assembly_fragment"
     misc_feature    26594..36585
                     /note="assembly_fragment"
     misc_feature    36686..43665
                     /note="assembly_fragment"
     misc_feature    43766..55870
                     /note="assembly_fragment"
     misc_feature    55970..66656
                     /note="assembly_fragment"
     misc_feature    66756..87066
                     /note="assembly_fragment"
     misc_feature    87167..109119
                     /note="assembly_fragment"
     misc_feature    109220..135181
                     /note="assembly_fragment"
     misc_feature    135282..160708
                     /note="assembly_fragment"
```

```
RESULT 7
LOCUS       AC097372                163823 bp    DNA     linear   PRI 01-MAR-2002
AC097372/c
DEFINITION  Homo sapiens BAC clone RP11-6L6 from 4, complete sequence.
ACCESSION   AC097372 AC010832
VERSION     AC097372.3  GI:16874901
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163823)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 163823)
  AUTHORS   Cedroni,M., Kozlowicz,A. and Elliott,G.
  TITLE     The sequence of Homo sapiens BAC clone RP11-6L6
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 163823)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 163823)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 163823)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 163823)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Department of Genetics, Washington
```

```
Qy          4 TTCTAGTCGACAAGAGTTTGTCAGTGGG 31
              |||||||| || ||||||||||||||||
Db 112934 TTCTAGTCCACATGATTTTGTCTGTGGG 112907
```

```
          Query Match        65.5%;  Score 21.6;  DB 2;  Length 160708;
          Best Local Similarity  85.7%;  Pred. No. 36;
          Matches   24; Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

BASE COUNT    48201 a   32652 c  31946 g  46605 t    1304 others
ORIGIN
```

```
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Nov 9, 2001 this sequence version replaced gi:16445193.
            ------------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: saplens@watson.wustl.edu
            ------------------- Summary Statistics
                Center project name: H_NH0005L06
                Drafting Center: WIBR
            -------------------

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR: pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-203B7; the clone sequenced
            to the right is RP11-552I10, 2000 bp overlap. Actual start of this
            clone is at base position 1 of RP11-6L6.

            Data from AC009875 was used to finish this clone, AC097372.

            The sequence of AC010832 has been incorporated into AC097372.
FEATURES             Location/Qualifiers
     source          1..163823
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-6L6"
                     /clone_lib="RPCI-11"
     repeat_region   21..50
                     /rpt_family="(TTTC)n"
     repeat_region   33..160
                     /rpt_family="Alu"
     repeat_region   452..759
                     /rpt_family="Alu"
     repeat_region   736..758
                     /rpt_family="(TAAA)n"
     repeat_region   759..793
                     /rpt_family="AT_rich"
     repeat_region   955..1010
                     /rpt_family="MER1_type"
     misc_feature    2344..2811
                     /note="match to EST AI821600 (NID:g5440679) nc46d05.x5"
     misc_feature    2344..2762
```

```
     misc_feature    2609..2792
                     /note="match to EST AA228735 (NID:g1851396) nc46d05.r1"
     repeat_region   2825..3467
                     /rpt_family="ERVL"
     repeat_region   4123..4287
                     /rpt_family="MIR"
     repeat_region   6799..6930
                     /rpt_family="MIR"
     repeat_region   7384..7405
                     /rpt_family="AT_rich"
     repeat_region   7699..8047
                     /rpt_family="MaLR"
     repeat_region   8343..8703
                     /rpt_family="Alu"
     repeat_region   9536..9595
                     /rpt_family="MER1_type"
     repeat_region   9609..9629
                     /rpt_family="(TA)n"
     repeat_region   9614..11418
                     /rpt_family="AT_rich"
     repeat_region   11422..11524
                     /rpt_family="L1"
     repeat_region   11525..11710
                     /rpt_family="L1"
     repeat_region   11711..12027
                     /rpt_family="MER1_type"
     repeat_region   11832..11852
                     /rpt_family="Alu"
     repeat_region   12011..12035
                     /rpt_family="AT_rich"
     repeat_region   12028..12072
                     /rpt_family="(TAAAA)n"
     repeat_region   12073..12151
                     /rpt_family="MER1_type"
     repeat_region   12191..12354
                     /rpt_family="L1"
     repeat_region   12276..12321
                     /rpt_family="MER1_type"
     repeat_region   12370..12731
                     /rpt_family="AT_rich"
     repeat_region   13182..13462
                     /rpt_family="L1"
     repeat_region   13463..13612
                     /rpt_family="Alu"
     repeat_region   13464..13583
                     /rpt_family="ERVL"
     repeat_region   13685..13877
                     /rpt_family="(GA)n"
     misc_feature    14321..14348
                     /rpt_family="MIR"
     repeat_region   14321..14343
                     /note="match to EST AI425053 (NID:g4270984) tg950g01.x1"
     misc_feature    14421..14341
                     /rpt_family="(T)n"
     repeat_region   15556..15861
                     /note="match to EST AI125330 (NID:g3593844) qd92e10.x1"
     repeat_region   16673..16762
                     /rpt_family="Alu"
     repeat_region   17608..17696
                     /rpt_family="MIR"
     repeat_region   18069..18134
                     /rpt_family="MIR"
     repeat_region   18162..18196
                     /rpt_family="MIR"
     misc_feature    18753..19167
                     /note="match to EST AI125330 (NID:g3593844) qd92e10.x1"
     misc_feature    18799..19167
                     /note="match to EST AI125266 (NID:g3593780) qd92a12.x1"
     repeat_region   18853..18983
                     /rpt_family="CR1"
     repeat_region   19065..19300
                     /rpt_family="MIR"
     repeat_region   19660..19698
                     /rpt_family="MIR"
```

```
repeat_region    /rpt_family="(CA)n"
                 1971..2005,
repeat_region    /rpt_family="L2"
                 2005..2013
repeat_region    /rpt_family="(CAAAA)n"
                 2123..20294
misc_feature     /note="match to EST AA682233 (NID:g2669365) ah5zb12.s1"
                 20123..20347
repeat_region    /rpt_family="MIR"
                 20152..20347
misc_feature     /note="match to EST AA781375 (NID:g2840706) aj25e07.s1"
                 20416..20417
misc_feature     /note="match to EST AA781375 (NID:g2840706) aj25e07.s1"
                 20460..20774
                 /note="match to EST AA781375 (NID:g2840706) aj25e07.s1"

Query Match        65.5%;       Score 21.6;  DB 9;   Length 163823;
Best Local Similarity   85.7%;       Pred. No. 36;
Matches   24;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

Qy      4 TTCTAGTCCACAAGACTTTCTCACTGGG 31
          |||||||||| ||||| ||||||||||||
Db  65387 TTCTAGTCCACATGATTTGTCTGTGGG 65360

RESULT 8
AC126835
LOCUS        AC126835             199006 bp    DNA     linear   HTG 10-JUL-2002
DEFINITION   Rattus norvegicus clone CH230-9B24, *** SEQUENCING IN PROGRESS ***,
             68 unordered pieces.
ACCESSION    AC126835
VERSION      AC126835.1  GI:21722693
KEYWORDS     HTG; HTGS_PHASE1.
SOURCE       Norway rat.
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1  (bases 1 to 199006)
  AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
             Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
             Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
             Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
             Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
             Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
             Cleveland,C.D., Cox,C., Coyle,M.D., Davila,R.,
             Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
             Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
             Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
             Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
             Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
             Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
             Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
             Harris,C., Harris,J., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
             Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
             Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
             Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
             Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
             Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
             Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
             Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
             Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
             Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okwuonu,G.,
             Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
             Peters,L., Pickens,R., Primus,E., Pu,L.L., Puiles,M., Ren,Y.,
             Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
             Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
             Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
             Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
             Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
             Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
             Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
             Williams,G., Williamson,A., Wieczyk,R., Wooden,S., Worley,K.,
             Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
             Weinstock,G. and Gibbs,R.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 199006)
  AUTHORS    Worley,K.C.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
             of Molecular and Human Genetics, Baylor College of Medicine, One
             Baylor Plaza, Houston, TX 77030, USA
COMMENT      Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
             ---------------------------------------------------
             Project Information
             Center project name: GDBV
             Center clone name: CH230-9B24
             ---------------------------------------------------
             Summary Statistics
             Sequencing vector: plasmid;
             Chemistry: Dye-terminator Big Dye: 100% of reads
             Assembly program: Phrap; version 0.990329
             Consensus quality: 139089 bases at least Q40
             Consensus quality: 145464 bases at least Q30
             Consensus quality: 150336 bases at least Q20
             ---------------------------------------------------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 68 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *
             *       1  1403: contig of 1403 bp in length
             *    1404  1503: gap of unknown length
             *    1504  3061: contig of 1558 bp in length
             *    3062  3161: gap of unknown length
             *    3162  4558: contig of 1397 bp in length
             *    4559  4658: gap of unknown length
             *    4659  6132: contig of 1474 bp in length
             *    6133  6232: gap of unknown length
             *    6233  7712: contig of 1480 bp in length
             *    7713  7812: gap of unknown length
             *    7813  8880: contig of 1068 bp in length
             *    8881  8980: gap of unknown length
             *    8981 10301: contig of 1321 bp in length
             *   10302 10401: gap of unknown length
             *   10402 11411: contig of 1010 bp in length
             *   11412 11511: gap of unknown length
             *   11512 12948: contig of 1437 bp in length
             *   12949 13048: gap of unknown length
             *   13049 14182: contig of 1134 bp in length
             *   14183 14282: gap of unknown length
             *   14283 14483: contig of 1178 bp in length
             *   14483 15560: gap of unknown length
             *   15561 15560: contig of 1245 bp in length
             *   15561 16805: gap of unknown length
             *   16806 16905: contig of 1321 bp in length
             *   16906 18428: gap of unknown length
             *   18429 18528: contig of 1523 bp in length
             *   18529 19864: gap of unknown length
             *   19865 19964: contig of 1336 bp in length
             *   19965 21006: gap of unknown length
             *   21007 21106: contig of 1042 bp in length
             *   21107 22206: gap of unknown length
             *   22207 22306: contig of 1100 bp in length
             *   22307 23657: gap of unknown length
             *   23658 23757: contig of 1351 bp in length
             *   23758 25200: gap of unknown length
             *   25200       : contig of 1443 bp in length
```

```
   25201   25300: gap of unknown length
   25301   25847: contig of 1547 bp in length
   26848   26947: gap of unknown length
   26948   28625: contig of 1678 bp in length
   28626   28725: gap of unknown length
   28726   30230: contig of 1505 bp in length
   30231   30330: gap of unknown length
   30331   31442: contig of 1112 bp in length
   31443   31542: gap of unknown length
   31543   32563: contig of 1021 bp in length
   32564   32663: gap of unknown length
   32664   34731: contig of 2068 bp in length
   34732   34831: gap of unknown length
   34832   36545: contig of 1714 bp in length
   36546   36645: gap of unknown length
   36646   37805: contig of 1160 bp in length
   37806   37905: gap of unknown length
   37906   39636: contig of 1731 bp in length
   39637   39736: gap of unknown length
   39737   41668: contig of 1932 bp in length
   41669   41768: gap of unknown length
   41769   43961: contig of 2193 bp in length
   43962   44061: gap of unknown length
   44062   46438: contig of 2377 bp in length
   46439   46538: gap of unknown length
   46539   47855: contig of 1317 bp in length
   47856   47955: gap of unknown length
   47956   49900: contig of 1945 bp in length
   49901   50000: gap of unknown length
   50001   51613: contig of 1613 bp in length
   51614   51713: gap of unknown length
   51714   54453: contig of 2740 bp in length
   54454   54553: gap of unknown length
   54554   56445: contig of 1892 bp in length
   56446   56545: gap of unknown length
   56546   58770: contig of 2225 bp in length
   58771   58870: gap of unknown length
   58871   61480: contig of 2610 bp in length
   61481   61580: gap of unknown length
   61581   64286: contig of 2706 bp in length
   64287   64386: gap of unknown length
   64387   66987: contig of 2601 bp in length
   66988   67087: gap of unknown length
   67088   69099: contig of 2012 bp in length
   69100   69199: gap of unknown length
   69200   71985: contig of 2786 bp in length
   71986   72085: gap of unknown length
   72086   73999: contig of 1914 bp in length
   74000   74099: gap of unknown length
   74100   76895: contig of 2796 bp in length
   76896   76995: gap of unknown length
   76996   79677: contig of 2682 bp in length
   79678   79777: gap of unknown length
   79778   82188: contig of 2411 bp in length
   82189   82288: gap of unknown length
   82289   84497: contig of 2209 bp in length
   84498   84597: gap of unknown length
   84598   88360: contig of 3763 bp in length
   88361   88460: gap of unknown length
   88461   92497: contig of 4037 bp in length
   92498   92597: gap of unknown length
   92598   96247: contig of 3650 bp in length
   96248   96347: gap of unknown length
   96348   99319: contig of 2972 bp in length
   99320   99419: gap of unknown length
   99420  102725: contig of 3306 bp in length
  102726  102825: gap of unknown length
  102826  106854: contig of 4029 bp in length
  106855  106954: gap of unknown length
  106955  110069: contig of 3115 bp in length
  110070  110169: gap of unknown length
  110170  112355: contig of 2186 bp in length
  112356  112455: gap of unknown length
```

```
* 112456  115811: contig of 3356 bp in length
* 115812  115911: gap of unknown length
* 115912  120810: contig of 4899 bp in length
* 120811  120910: gap of unknown length
* 120911  126265: contig of 5359 bp in length
* 126266  126369: gap of unknown length
* 126270  126369: gap of unknown length
```

```
Query Match    64.8%;  Score 21.4;  DB 2;  Length 199006;
Best Local Similarity   80.6%;  Pred. No. 44;
Matches   25;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

Qy      2 CATTCTTAGTCGACAAGAGTTTGTCACTGGGA 32
          |||||| ||| ||||| ||||||| ||| ||||
Db  22157 CATTCTTGTCAACAGGAGGTTTCTCACTCGGA 22187
```

```
RESULT 9
AC122578/c
LOCUS       AC122578              161308 bp    DNA     linear   HTG 12-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-149L3, *** SEQUENCING IN PROGRESS
            ***, 76 unordered pieces.
ACCESSION   AC122578
VERSION     AC122578.2  GI:21728453
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 161308)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,S., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newson,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oraguwe,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,M., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161308)
```

```
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (25-MAY-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 161308)
AUTHORS   Worley,K.C.
TITLE     Direct Submission
JOURNAL   Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
          On Jul 11, 2002 this sequence version replaced gi:21205895.
COMMENT
          -------------- Genome Center
          Center: Baylor College of Medicine
          Center code: BCM
          Web site: http://www.hgsc.bcm.tmc.edu/
          Contact: hgsc-help@bcm.tmc.edu
          -------------- Project Information
          Center project name: GHCC
          Center clone name: CH230-149L3
          -------------- Summary Statistics
          Sequencing vector: plasmid;
          Chemistry: Dye-terminator Big Dye: 100% of reads
          Assembly program: Phrap; version 0.990329
          Consensus quality: 98694 bases at least Q40
          Consensus quality: 104469 bases at least Q30
          Consensus quality: 108317 bases at least Q20
          --------------
        * NOTE: Estimated insert size may differ from sequence length
        * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 76 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *
        *          1          1324: contig of 1324 bp in length
        *       1325          1424: gap of unknown length
        *       1425          2840: contig of 1416 bp in length
        *       2841          2940: gap of unknown length
        *       2941          4430: contig of 1490 bp in length
        *       4431          4530: gap of unknown length
        *       4531          5562: contig of 1032 bp in length
        *       5563          5662: gap of unknown length
        *       5663          6677: contig of 1015 bp in length
        *       6678          6777: gap of unknown length
        *       6778          8046: contig of 1269 bp in length
        *       8047          8146: gap of unknown length
        *       8147          9215: contig of 1069 bp in length
        *       9216          9315: gap of unknown length
        *       9316         10537: contig of 1222 bp in length
        *      10538         10637: gap of unknown length
        *      10638         11753: contig of 1116 bp in length
        *      11754         11853: gap of unknown length
        *      11854         12873: contig of 1020 bp in length
        *      12874         12973: gap of unknown length
        *      12974         14309: contig of 1336 bp in length
        *      14310         14409: gap of unknown length
        *      14410         15742: contig of 1333 bp in length
        *      15743         15842: gap of unknown length
        *      15843         17443: contig of 1601 bp in length
        *      17444         17543: gap of unknown length
        *      17544         18981: contig of 1438 bp in length
        *      18982         19081: gap of unknown length
        *      19082         20458: contig of 1377 bp in length
        *      20459         20558: gap of unknown length
        *      20559         22160: contig of 1602 bp in length
        *      22161         22260: gap of unknown length
        *      22261         23672: contig of 1412 bp in length
        *      23673         23772: gap of unknown length
        *      23773         25092: contig of 1320 bp in length
        *      25093         25192: gap of unknown length
        *      25193         26875: contig of 1683 bp in length
        *      26876         26975: gap of unknown length
        *      26976         28234: contig of 1259 bp in length
        *      28235         28334: gap of unknown length
        *      28335         29972: contig of 1638 bp in length
        *      29973         30072: gap of unknown length
        *      30073         31396: contig of 1324 bp in length
        *      31397         31496: gap of unknown length
        *      31497         33026: contig of 1530 bp in length
        *      33027         33126: gap of unknown length
        *      33127         34367: contig of 1241 bp in length
        *      34368         34467: gap of unknown length
        *      34468         36105: contig of 1638 bp in length
        *      36106         36205: gap of unknown length
        *      36206         38371: contig of 2166 bp in length
        *      38372         38471: gap of unknown length
        *      38472         40222: contig of 1751 bp in length
        *      40223         40322: gap of unknown length
        *      40323         41814: contig of 1492 bp in length
        *      41815         41914: gap of unknown length
        *      41915         43199: contig of 1285 bp in length
        *      43200         43299: gap of unknown length
        *      43300         45421: contig of 2122 bp in length
        *      45422         45521: gap of unknown length
        *      45522         47208: contig of 1687 bp in length
        *      47209         47308: gap of unknown length
        *      47309         48381: contig of 1073 bp in length
        *      48382         48481: gap of unknown length
        *      48482         49753: contig of 1272 bp in length
        *      49754         49853: gap of unknown length
        *      49854         51476: contig of 1623 bp in length
        *      51477         51576: gap of unknown length
        *      51577         54753: contig of 3177 bp in length
        *      54754         54853: gap of unknown length
        *      54854         56777: contig of 1924 bp in length
        *      56778         56877: gap of unknown length
        *      56878         58008: contig of 1131 bp in length
        *      58009         58108: gap of unknown length
        *      58109         60433: contig of 2325 bp in length
        *      60434         60533: gap of unknown length
        *      60534         62301: contig of 1768 bp in length
        *      62302         62401: gap of unknown length
        *      62402         64051: contig of 1650 bp in length
        *      64052         64151: gap of unknown length
        *      64152         65590: contig of 1439 bp in length
        *      65591         65690: gap of unknown length
        *      65691         66695: contig of 1005 bp in length
        *      66696         66795: gap of unknown length
        *      66796         68000: contig of 1205 bp in length
        *      68001         68100: gap of unknown length
        *      68101         69681: contig of 1581 bp in length
        *      69682         69781: gap of unknown length
        *      69782         71509: contig of 1728 bp in length
        *      71510         71609: gap of unknown length
        *      71610         72940: contig of 1331 bp in length
        *      72941         73040: gap of unknown length
        *      73041         74410: contig of 1370 bp in length
        *      74411         74510: gap of unknown length
        *      74511         77256: contig of 2746 bp in length
        *      77257         77356: gap of unknown length
        *      77357         79692: contig of 2336 bp in length
        *      79693         79792: gap of unknown length
        *      79793         81845: contig of 2053 bp in length
        *      81846         81945: gap of unknown length
        *      81946         83946: contig of 2001 bp in length
        *      83947         84046: gap of unknown length
        *      84047         84747: contig of 701 bp in length
        *      84748         84847: gap of unknown length
        *      84848         86548: contig of 1701 bp in length
        *      86549         86648: gap of unknown length
        *      86649         88726: contig of 2078 bp in length
        *      88727         88826: gap of unknown length
        *      88827         91454: contig of 2628 bp in length
```

```
Query Match    Best Local Similarity   63.0%;   Score 20.8;  DB 2;  Length 161308;
Matches  25;   Conservative   78.1%;   Pred. No. 84;
                            0;   Mismatches    7;   Indels   0;   Gaps   0;

Qry    1 GCATTCTAGTCGACAAGAGTTTGTCACTGGGA 32
         ||.|||.|||.||||.||||||.||.|||||||
Db 143193 GCTTTCCAGTAGACAAAAGAATGGCACTGGGA 143162
```

RESULT 10
ACl28864

LOCUS   ACl28864   179250 bp   DNA   linear   HTG 02-AUG-2002
DEFINITION Rattus norvegicus clone CH230-422P24, *** SEQUENCING IN PROGRESS
            ***, 55 unordered pieces.
ACCESSION   ACl28864
VERSION   ACl28864.2 GI:22038232
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Norway rat.
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 179250)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C.C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,F., Martinado,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,J., Nelson,K., Newton,J., Nickerson,E., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L., Puazo,M., Quiles,M.M.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,C., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 179250)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 179250)
  AUTHORS   Worley,K.C.

COMMENT   On Aug 1, 2002 this sequence version replaced gi:21953007.
            -------------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------------- Project Information
               Center project name: KAWO
               Center clone name: CH230-422P24
            -------------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 127967 bases at least Q40
               Consensus quality: 135422 bases at least Q30
               Consensus quality: 140879 bases at least Q20
            ----
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 55 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1: 1509: contig of 1509 bp in length
            *   1510: 1609: gap of unknown length
            *   1610: 2783: contig of 1174 bp in length
            *   2784: 2883: gap of unknown length
            *   2884: 3895: contig of 1012 bp in length
            *   3896: 3995: gap of unknown length
            *   3996: 5157: contig of 1162 bp in length
            *   5158: 5257: gap of unknown length
            *   5258: 6448: contig of 1191 bp in length
            *   6449: 6548: gap of unknown length
            *   6549: 7719: contig of 1171 bp in length
            *   7720: 7819: gap of unknown length
            *   7820: 8882: contig of 1063 bp in length
            *   8883: 8982: gap of unknown length
            *   8983: 10855: contig of 1873 bp in length
            *   10856: 10955: gap of unknown length
            *   10956: 12045: contig of 1090 bp in length
            *   12046: 12145: gap of unknown length
            *   12146: 13327: contig of 1182 bp in length
            *   13328: 13427: gap of unknown length
            *   13428: 14604: contig of 1177 bp in length
            *   14605: 14704: gap of unknown length
            *   14705: 15815: contig of 1111 bp in length
            *   15816: 15915: gap of unknown length
            *   15916: 17158: contig of 1243 bp in length
            *   17159: 17258: gap of unknown length
            *   17259: 18348: contig of 1090 bp in length
            *   18349: 18448: gap of unknown length
            *   18449: 19561: contig of 1113 bp in length
            *   19562: 19661: gap of unknown length
            *   19662: 21326: contig of 1665 bp in length
            *   21327: 21426: gap of unknown length
            *   21427: 22575: contig of 1149 bp in length
            *   22576: 22675: gap of unknown length
            *   22676: 25291: contig of 2616 bp in length
            *   25292: 25391: gap of unknown length
            *   25392: 27995: contig of 2604 bp in length
            *   27996: 28095: gap of unknown length
            *   28096: 30167: contig of 2072 bp in length
            *   30168: 30267: gap of unknown length
            *   30268: 32001: contig of 1734 bp in length
            *   32002: 32101: gap of unknown length
```

```
RESULT 11
LOCUS       AC117985              180035 bp    DNA     linear   HTG 12-APR-2002
DEFINITION  Papio cynocephalus anubis clone RP41-415A2, WORKING DRAFT SEQUENCE,
            5 unordered pieces.
ACCESSION   AC117985
VERSION     AC117985.1  GI:20143536
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Papio cynocephalus anubis.
  ORGANISM  Papio cynocephalus anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 180035)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S-L., Idol,J.R., Karlins,E., Lario,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Maslello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180035)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     -----------------  Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            -----------------  Project Information
            Center project name: cqu
            Center clone name: 415A02
            -----------------  Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 178328 bases at least Q40
            Consensus quality: 178649 bases at least Q30
            Consensus quality: 178855 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 179635; sum-of-contigs
            Quality coverage: 9.97x in Q20 bases; agarose-fp
            Quality coverage: 10.10x in Q20 bases; sum-of-contigs
            -----------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1..180035
                     /organism="Papio cynocephalus anubis"
```

```
*      1     21961: contig of 21961 bp in length
*  21962     22061: gap of unknown length
*  22062     41049: contig of 18988 bp in length
*  41050     41150: gap of unknown length
*  41151     72888: contig of 31738 bp in length
*  72889     72988: gap of unknown length
*  72989    119485: contig of 46497 bp in length
* 119486    119585: gap of unknown length
* 119586    180035: contig of 60450 bp in length
```

```
*  32102     34134: contig of 2033 bp in length
*  34135     34234: gap of unknown length
*  34235     36533: contig of 2299 bp in length
*  36534     36633: gap of unknown length
*  36634     38118: contig of 1485 bp in length
*  38119     38218: gap of unknown length
*  38219     40043: contig of 1825 bp in length
*  40044     40143: gap of unknown length
*  40144     42111: contig of 1968 bp in length
*  42112     42211: gap of unknown length
*  42212     44790: contig of 2579 bp in length
*  44791     44890: gap of unknown length
*  44891     46831: contig of 1941 bp in length
*  46832     46931: gap of unknown length
*  46932     49886: contig of 2955 bp in length
*  49887     49986: gap of unknown length
*  49987     53806: contig of 3820 bp in length
*  53807     53906: gap of unknown length
*  53907     56935: contig of 3029 bp in length
*  56936     57035: gap of unknown length
*  57036     59283: contig of 2248 bp in length
*  59284     59383: gap of unknown length
*  59384     61818: contig of 2435 bp in length
*  61819     61918: gap of unknown length
*  61919     65811: contig of 3893 bp in length
*  65812     65911: gap of unknown length
*  65912     68623: contig of 2712 bp in length
*  68624     68723: gap of unknown length
*  68724     71718: contig of 2996 bp in length
*  71719     71819: gap of unknown length
*  71820     76222: contig of 4403 bp in length
*  76223     76322: gap of unknown length
*  76323     80018: contig of 3696 bp in length
*  80019     80118: gap of unknown length
*  80119     83971: contig of 3853 bp in length
*  83972     84071: gap of unknown length
*  84072     87959: contig of 3888 bp in length
*  87960     88059: gap of unknown length
*  88060     91801: contig of 3742 bp in length
*  91802     91901: gap of unknown length
*  91902     97354: contig of 5453 bp in length
*  97355     97454: gap of unknown length
*  97455    100353: contig of 2899 bp in length
* 100354    100453: gap of unknown length
* 100454    105784: contig of 5331 bp in length
* 105785    105884: gap of unknown length
* 105885    110762: contig of 4878 bp in length
* 110763    110862: gap of unknown length
* 110863    114911: contig of 4049 bp in length
* 114912    115011: gap of unknown length
* 115012    120392: contig of 5381 bp in length
* 120393    120492: gap of unknown length
* 120493    126236: contig of 5744 bp in length
* 126237    126336: gap of unknown length
* 126337    132710: contig of 6374 bp in length
* 132711    132810: gap of unknown length
* 132811    139090: contig of 6280 bp in length
* 139091    139190: gap of unknown length
* 139191    143974: contig of 4784 bp in length
* 143975    144074: gap of unknown length
* 144075    148570: contig of 4496 bp in length
* 148571    148670: gap of unknown length
* 148671    153393: contig of 4723 bp in length
* 153394    153493: gap of unknown length
* 153494    159140: contig of 5647 bp in length
```

```
Query Match             63.0%;  Score 20.8;  DB 2;  Length 179250;
Best Local Similarity   78.1%;
Matches   25;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;
                        Pred. No. 83;

Qy        1 GCATTCTAGTCGGACAAAGAGTTTGTCAGTGGGA 32
            || |||  || || |||||||| | |||||||||
Db   129473 GCTTTCCAGTAGACAAAAGAATGGCAGTGGGA 129504
```

misc_feature
    /db_xref="taxon:9555"
    /clone="RP41-415A2"
    /note="assembly_fragment"
misc_feature
    22062..41050
    /note="assembly_fragment"
    /clone_lib="RP41"
    /clone_end:T7
    vector_side:right"
misc_feature
    41151..72888
    /note="assembly_fragment"
misc_feature
    72989..119485
    /note="assembly_fragment"
misc_feature
    119586..180035
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left"

ORIGIN
BASE COUNT   53465 a   38533 c   37931 g   49705 t   401 others

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGAG 33
         |||| | | ||||||||| ||||||||||||||
Db 80872 CATTTTGTTACAGAAGAGTTTATCAGTGGGAG 80903

RESULT 12
LOCUS   AC116932
ACCESSION   AC116932
VERSION   AC116932.1 GI:19909406
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Papio cynocephalus anubis.
ORGANISM   Papio cynocephalus anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1 (bases 1 to 180783)
AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
            Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E., Laric,P.,
            Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., Mastrian,S.D.,
            McCloskey,J.C., McDowell,J., Pagulirigan,C., Pearson,R.,
            Portnoy,M.E., Prasad,A., Schueler,M.G., Stantripop,S., Thomas,J.W.,
            Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
TITLE   NISC Comparative Sequencing Initiative
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 180783)
AUTHORS   Green,E.D.
TITLE   Direct Submission
JOURNAL   Submitted (03-APR-2002) NIH Intramural Sequencing Center,
            Gaithersburg, MD 20877, USA
COMMENT   -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
            Center project name: cqt
            Center clone name: 156H04
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly Program: Phrap; version 0.990319
            Consensus quality: 180093 bases at least Q40

LOCUS   AC116932   180783 bp   DNA   linear   HTG 03-APR-2002
DEFINITION Papio cynocephalus anubis clone RP41-156H4, WORKING DRAFT SEQUENCE,
            2 unordered pieces.

Consensus quality: 180363 bases at least Q30
Consensus quality: 180474 bases at least Q20
Insert size: 124000; agarose-fp
Insert size: 180683; sum-of-contigs
Quality coverage: 12.22x in Q20 bases; agarose-fp
Quality coverage: 8.38x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*   1   90657: contig of 90657 bp in length
*   90658   90757: gap of unknown length
*   90758   180783: contig of 90026 bp in length

FEATURES   Location/Qualifiers
     source   1..180783
                     /organism="Papio cynocephalus anubis"
                     /db_xref="taxon:9555"
                     /clone="RP41-156H4"
     misc_feature   1..90657
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature   90758..180783
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"

ORIGIN
BASE COUNT   53019 a   40040 c   40244 g   47368 t   112 others

Qy   2 CATTCTAGTCGACAAGAGTTTGTCAGTGGGAG 33
         |||| | | ||||||||| ||||||||||||||
Db 90608 CATTTTGTTACAGAAGAGTTTATCAGTGGGAG 90639

RESULT 13
LOCUS   AL606464/C
ACCESSION   AL606464
VERSION   AL606464.11 GI:16073721
KEYWORDS   HTG.
SOURCE   Mus musculus
ORGANISM   Mus musculus
            house mouse.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 183698)
AUTHORS   Bates,K.
TITLE   Direct Submission
JOURNAL   Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Oct 11, 2001 this sequence version replaced gi:16030239.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such

LOCUS   AL606464   183698 bp   DNA   linear   ROD 29-JUN-2002
DEFINITION Mouse DNA sequence from clone RP23-462P2 on chromosome 13, complete
            sequence.

```
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-462P2 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

FEATURES             Location/Qualifiers
     source          1..183698
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-462P2"
                     /clone_lib="RPCI-23"
BASE COUNT     53761 a   39743 c   39566 g   50628 t
ORIGIN

Query Match             63.0%;  Score 20.8;  DB 10;  Length 183698;
  Best Local Similarity 78.1%;  Pred. No. 83;
  Matches  25;  Conservative  0;  Mismatches  7;  Indels  0;  Gaps  0;

Qy      1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
          ||||||||| |||| ||||||| |||||||| ||
Db  66441 GCATTCTAGTCACCAAGCCTTTGTCAATAAGA 66410

RESULT 14
LOCUS       AL627315/c        209876 bp    DNA     linear   HTG 19-JUN-2002
DEFINITION  Mus musculus chromosome 13 clone RP23-212D7, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL627315
VERSION     AL627315.7  GI:18135166
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            house mouse
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209876)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 11, 2002 this sequence version replaced gi:17048468.
            -------------------- Genome Center --------------------
            Center: UK Medical Research Council
            Center code: UK-MRC
            Web site: http://mrcseq.har.mrc.ac.uk
            Contact: mouseq@har.mrc.ac.uk
            ----------------------------------------
            -------- Project Information -----------
            Center project name: bm212D7
            ----------------------------------------
            --------- Summary Statistics -----------
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 209092 bases at least Q40
            Consensus quality: 209289 bases at least Q30
            Consensus quality: 209389 bases at least Q20
            Insert size: 209876; sum-of-contigs
            Insert size: 204253; 11.7% error; agarose-fp
            Quality coverage: 9.16x in Q20 bases; sum-of-contigs Quality
```

```
coverage: 9.41x in Q20 bases; agarose-fp
            *  NOTE: This is a 'working draft' sequence.
            *  This sequence will be replaced
            *  by the finished sequence as soon as it is available and
            *  the accession number will be preserved.

FEATURES             Location/Qualifiers
     source          1..209876
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="13"
                     /clone="RP23-212D7"
                     /clone_lib="RPCI-23"
     misc_feature    1..209876
                     /note="assembly_fragment:04602"
BASE COUNT     60767 a   43745 c   44655 g   60709 t
ORIGIN

Query Match             63.0%;  Score 20.8;  DB 2;  Length 209876;
  Best Local Similarity 78.1%;  Pred. No. 82;
  Matches  25;  Conservative  0;  Mismatches  7;  Indels  0;  Gaps  0;

Qy      1 GCATTCTAGTCGACAAGAGTTTGTCAGTGGGA 32
          ||||||||| |||| ||||||| |||||||| ||
Db 190393 GCATTCTAGTCACCAAGCCTTTGTCAATAAGA 190362

RESULT 15
LOCUS       AC122449/c        169514 bp    DNA     linear   HTG 05-JUL-2002
DEFINITION  Mus musculus chromosome UNK clone RP24-247014, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC122449
VERSION     AC122449.2  GI:21699710
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            house mouse
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 169514)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     The sequence of Mus musculus clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169514)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   3  (bases 1 to 169514)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
COMMENT     On Jul 5, 2002 this sequence version replaced gi:21105910.
            -------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ----------------------------------------
            -------- Project Information -----------
            Center project name: M_BB0247014
            ----------------------------------------
            --------- Summary Statistics -----------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 166352 bases at least Q40
            Consensus quality: 166800 bases at least Q30
            Consensus quality: 167046 bases at least Q20
            Insert size: 169k; agarose-fp
```

```
           Insert size: 168914; sum-of-contigs
           Quality coverage: 0.00 in Q20 bases; agarose-fp
           Quality coverage: 11.90 in Q20 bases; sum-of-contig

      * NOTE: This is a 'working draft' sequence. It currently
      * consists of 7 contigs. The true order of the pieces
      * is not known and their order in this sequence record is
      * arbitrary. Gaps between the contigs are represented as
      * runs of N, but the exact sizes of the gaps are unknown.
      * This record will be updated with the finished sequence
      * as soon as it is available and the accession number will
      * be preserved.
      *
      *    1            1338: contig of 1338 bp in length
      *    1339         1438: gap of unknown length
      *    1439         6519: contig of 5081 bp in length
      *    6520         6519: gap of unknown length
      *    6520        17199: contig of 10580 bp in length
      *   17200        17299: gap of unknown length
      *   17300        26065: contig of 8766 bp in length
      *   26066        26165: gap of unknown length
      *   26166        82538: contig of 56373 bp in length
      *   82539        82638: gap of unknown length
      *   82639       168110: contig of 85472 bp in length
      *  168111       168210: gap of unknown length
      *  168211       169514: contig of 1304 bp in length.

FEATURES           Location/Qualifiers
     source         1..169514
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /chromosome="UNK"
                    /clone="RP24-247014"
     misc_feature   1..1338
                    /note="assembly_name:Contig13"
     misc_feature   1439..6519
                    /note="assembly_name:Contig14"
     misc_feature   6520..17199
                    /note="assembly_name:Contig15"
     misc_feature   17300..26065
                    /note="assembly_name:Contig16"
     misc_feature   26166..82538
                    /note="assembly_name:Contig17"
     misc_feature   82639..168110
                    /note="assembly_name:Contig18"
     misc_feature   168211..169514
                    /note="assembly_name:Contig19"
BASE COUNT     44433 a   39379 c   41318 g   43780 t    604 others
ORIGIN


Query Match        62.4%;   Score 20.6;   DB 2;   Length 169514;
Best Local Similarity   85.2%;   Pred. No. 1e+02;
Matches   23;   Conservative    0;   Mismatches    4;   Indels    0;   Gaps    0;

Qy       3 ATTTCTAGTCGACAAGAGTTTGTCAGTG 29
           |||||||| | ||||||| | |||||||| |
Db   52765 ATTTCTAGCCAACAAGAATTTGTCAGG 52739

Search completed: December 16, 2002, 12:43:33
Job time : 3614 secs
```